DOCUMENT RESUME

ED 372 717 HE 027 672

AUTHOR	Greenwood, Addison
TITLE	The National Assessment of College Student Learning: Identification of the Skills To Be Taught, Learned, and Assessed. A Report on the Proceedings of the Study Design Workshop (2nd, Arlington, Virginia, November 17-18, 1992). Research and Development Report.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	ISBN-0-16-045146-9; NCES-94-286
PUB DATE	Aug 94
NOTE	321p.; For related reports, see HE 027 673, ED 363 177, ED 359 856 and ED 346 808.
AVAILABLE FROM	U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	Citizenship; *College Outcomes Assessment; Communication Skills; *Critical Thinking; Delphi Technique; *Educational Assessment; *Educational Objectives; Higher Education; Listening Skills; *Problem Solving; Reading Skills; Speech Skills; Standards; Thinking Skills; Writing Skills
IDENTIFIERS	*Indicators; *National Education Goals 1990

ABSTRACT

This report presents the primary papers given at a 1992 workshop to identify specific higher order thinking and communication skills and to develop appropriate indicators of collegiate outcomes as suggested by Goal 5.5 of the National Education Goals formulated in 1990. This goal addressed improvement in college graduates' ability to "think critically, communicate effectively, and solve problems." Part 1 presents the five papers upon which the working groups built their discussions. These papers are: (1) "Skills for Citizenship" (Suzanne W. Morse); (2) "A National Assessment of Critical Thinking Skills in Adults: Taking Steps Toward the Goal" (Diane F. Halpern); (3) "Assessing Thinking: A Framework for Measuring Critical Thinking and Problem Solving Skills at the College Level" (David Perkins et al.); (4) "Assessing Speaking and Listening: Preliminary Considerations for a National Assessment" (John A. Daly); and (5) "No Guru, No Method, No Teacher: The Communication Domain and the NACSL (National Assessment of College Student Learning)" by Stephen P. Witte. Part 2 includes group summary reports, reviewers' comments, a listing of speaking and listening skills, and additional information on an ongoing Delphi study at Pennsylvania State University. Part 3 presents comments by participants on the conference products and results of a survey of all participants. Consensus on "next steps" included recommending further development of the taxonomy of skills, abilities, and competencies. (The five papers contain references.) (DB)

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: HE

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view TM

ED 372 717

NATIONAL CENTER FOR EDUCATION STATISTICS

Research and Development Report August 1994

The National Assessment of College Student Learning: Identification of the Skills to Be Taught, Learned, and Assessed

A Report on the Proceedings of the Second Study Design Workshop November 1992

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE027672

U.S. Department of Education
Office of Educational Research and Improvement NCES 94-286

BEST COPY AVAILABLE

NATIONAL CENTER FOR EDUCATION STATISTICS

Research and Development Report August 1994

The National Assessment of College Student Learning: Identification of the Skills to Be Taught, Learned, and Assessed

A Report on the Proceedings of the Second Study Design Workshop November 1992

Addison Greenwood
Editor
Washington, DC

Sal Corrallo, Project Director
Educational Assessment Division
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement NCES 94-286

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Commissioner

National Center for Education Statistics

"The purpose of the Center shall be to collect, analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e–1).

August 1994

Contact:
Sal Corrallo
(202) 219–1913

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-045146-9

FOREWORD

This is the third in a series of working papers, published by the National Center for Education Statistics, on the assessment of college student learning. The work has been conducted in direct support of National Education Goal 5:

> "By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship."

Specifically this third workshop focused on Objective 5:

> "The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially."

Together with the earlier working papers, the work reported here provides a solid base of knowledge on which to stand in the effort to formally identify workplace and citizenship skills. This work shows that the concerns of several varied publics must be considered and balanced in an open and fair forum.

This working paper, like the first two publications, was born primarily from the ideas and opinions of the authors, reviewers, and workshop participants. We are grateful for their willingness to share their time and ideas with us. Also a tribute goes to Addison Greenwood, who summarized the proceedings of both workshops and synthesized the commissioned papers in the first workshop. We are grateful to him for his effort to capture and report the ideas and concerns of the authors and workshop participants in summary form, fairly yet with the necessary detail to ensure their integrity. A special tribute also to the facilitators and recorders (named in the workshop sections of this paper), to Rita Talfero of Associate Consultants, and to other support staff for their untiring contributions. Finally our thanks to the NCES project staff headed by Sal Corrallo and assisted by David Loope and Angie Miles, for their guidance and management of the project.

Gary Phillips,
Associate Commissioner, EAD/NCES

Emerson Elliott,
Commissioner, NCES

TABLE OF CONTENTS

FOREWORD iii

Part 1. PREPARATION

Introduction 1
Contents of this Paper 8
Keynote Remarks 9

Skills for Citizenship, ***Suzanne W. Morse*** 14

Informing Their Discretion 15
What Citizens Need To Know 15
Role of Higher Education 20
References 22

A National Assessment of Critical Thinking Skills in Adults: Taking Steps Toward the Goal, ***Diane F. Halpern*** 24

The Case for a National Assessment of Critical Thinking Skills in Adults 24
    Improving Teaching and Learning 25
    Accountability 26
    The Crisis in Human Capital 27
    Creating Self-Fulfilling Prophecies 28

What Are Critical Thinking Skills? 28
    Rationale and Criteria for Targeted Skills 28
    A Taxonomy of Critical Thinking Skills 31
    College-Level Critical Thinking Skills 33
    The Disposition To Think Critically 35
    Evidence That Critical Thinking Skills Can Be Learned 36

Resolving Critical Issues in Test Design and Administration 38
    Validity 38
    Unit of Analysis 40
    Test and Question Format 41
    Secondary Indicators 45
    Standards 47
    Communicating Results 48

A Plan for Test Development 49
    Staged Research and Development 49
    Administration of Pilot Instrument 50

Simultaneous Administration of Skills Surveys 51
What We Can Learn From This Assessment 51

Caveats and Conclusions 52
Resistance and Realistic Expectations 53

References 54
Appendices
Appendix A. Descriptions and Examples From the College-Level Thinking Skills Taxonomy 57
Appendix B. Sample Questions 63

Assessing Thinking: A Framework for Measuring Critical-Thinking and Problem-Solving Skills at the College Level, ***David Perkins, Eileen Jay, and Shari Tishman*** . . 65

Framing the Mission 65
The Plan of the Paper 66
Fly-on-the-wall Assessment 67
The Importance of Dispositions Alongside Abilities 68
The Importance of Kinds of Thinking 69

A Framework for Assessment 70
Defining an Approach 70
Seven Key Dispositions and Abilities 71
Basic Problem Types 73
Features of the Larger Setting 76
How Assessment Might Work 77

Rationale for the Framework: Six Useful Features 78
Feature 1: A More Comprehensive, Two-part Framework of Thinking: Dispositions and Abilities and Basic Problem Types 79
Feature 2: A Manageable Number of Broad Encompassing Categories 81
Feature 3: Categories That Work Across Different Contexts of Knowledge 82
Feature 4: Dispositions Defined in Behavioral Terms for Easy Identification 83
Feature 5: Dispositions and Abilities Paired for Structural Simplicity 84
Feature 6: Greater Integration of Standards into Dispositions, Abilities, and Basic Problem Types to Reflect Their Central Role. 86

Three Principles for an Assessment Methodology 88
Use Authentic Assessments 88
Probe Thinking Processes, Not Just Products 89
Sample Different Contexts and Settings of Knowledge 90
Teaching to Tests Worth Teaching to: An Outlook for Education 92
References 94

Appendix: Sample Assessments 98

Assessing Speaking and Listening: Preliminary Considerations for a National Assessment, *John A. Daly* 113

The Importance of Communication 113
Basic Conceptual Concerns 114
Issues in the Assessment of Communication Skills 116
Assessing Communication Skills: What Should We Assess? 121
  Assessing Motivations and Predispositions to Communicate 121
  Assessing Knowledge 122
  Assessing Performance 123
  Public Speaking Assessment 123
  Assessing Interviewing Skills 125
  Listening Assessment 126
  Assessing Conversational Performance 128
  Assessing Small Group Exchanges 130
Assessment Criteria 132
Functions of Communication 133
References 135
Appendices 141

No Guru, No Method, No Teacher: The Communication Domain and the NACSL, *Stephen P. Witte* 181
The Communication Domain: Writing and Reading 182
  Some Theoretical Issues Practically Situated 183
  Practicing Theory, Theorizing Practice: An Example 189
  From Theorizing and Practicing to Skills for Assessing 194
  Through Indirection Finding Directions Out: Skills, Performances, Levels . . . 204
Notes Toward a Communication Assessment Model 208
Some After Words 211
References 212
Supplement, Response to Daly and Witte Papers, *Marcia Farr* 220

Part 2. THE WORKING GROUPS

Introduction to Working Group Reports 227
Problem Solving and Critical Thinking, Group 1 229
Problem Solving and Critical Thinking, Group 2 251
Speaking and Listening, Group 3 269
Reading and Writing, Group 4 282

Part 3. CLOSING COMMENTS

Final Open Session-Individual Comments 295
Where To Go From Here 311
1993 Activities 313

Part 1. PREPARATION

INTRODUCTION

In conversations about college learning throughout America, one hears the deeply felt and diverse concerns of educators, employers, and the public at large. Objective 5 of Goal 5 provides the National Center for Education Statistics (NCES) a unique opportunity to transform this national discussion into a collaborative exercise where these concerns can be effectively addressed.

Can these interested though diverse publics agree on a valid system to assess the higher order thinking and communication skills of college graduates? The answer to this larger question entails a more practical, immediate one: *How* can the various interests and publics come together in a working effort to develop such a framework? NCES has begun to answer this practical question by bringing representatives of these publics into a dialogue, facilitating the collection of ideas, and encouraging them to propose answers to what have emerged as the core questions: What skills should be targeted, what standards established, and how can we ensure that the assessment effort will improve teaching and learning?

The purpose of this publication and the two that preceeded it is to publish and broadcast that discussion, to memorialize the ideas prompted by these core questions. Developing a system where teaching and learning are enhanced in the very process of assessing the designated skills, relative to the emergent standards, will be no easy task. A National Goals Panel report released in July 1992 noted that

> "the Task Force recognizes that the development of national standards and assessment measures for American postsecondary education is a long-term effort requiring much consensus building, assessment development, and testing. It concludes, however, that the nations's future well-being depends upon the establishment of standards for learning which will raise educational achievement levels sufficient to insure that the United States' citizens are competitive with their counterparts throughout the world. The development of national standards of performance and assessment requires national leadership and federal support."

The report carried two further recommendations, that:

1. "The Goals Panel encourage the development of a sample-based national system of standards and assessment for postsecondary education"

2. "The Goals Panel suggest that content and performance standards be developed for general cognitive skills, higher order thinking skills, and occupational specific skills where appropriate."[1]

In the ideal, meeting Goal 5.5 could serve twin interests. Social and economic results should accrue to society when the educational process can deliver citizens capable of better exercising both power and responsibility. And this improved citizenry will arise from the development of individual students, who will reap lifelong personal benefits by investing their time and energy in a system designed around skills and standards that underlie subsequent success. The enterprise will fortify the link between the traditional mission of postsecondary institutions--to produce an enlightened, educated person--and the needs of industry for a capable and effective workforce.

The 1991 Work

In the effort to organize and encourage this conversation, NCES has thus far sponsored two workshops, in November of 1991 and 1992, both held in Washington, D.C.[2] The current document portrays the results of the 1992 two-day workshop, where the primary purpose was to identify specific higher order thinking and communication skills, and to develop a schema for distinguishing ability levels as suggested by Goal 5.5. The several score of scholars and educators who undertook this task did so in the context developed during the 1991 workshop. Were it not for the effort of nearly a hundred scholars who convened for that conference (only a few of whom were also part of the 1992 workshop), the considerable progress reflected here could not have been made, for a number of crucial issues had been raised and clarified, if not laid to rest. Among these issues were the following:

○ The original Goals Panel distinction between problem solving and critical thinking could be misleading. While academic distinctions of the study of problem solving persist, in the postsecondary world problem solving is the major *context* in which critical thinking skills are applied. These critical thinking skills and some of the mainline issues surrounding them can now be considered in the light of more than a decade of respectable academic work in that field.

○ The phrase "effective communication" (also inherited from the original goals) masks an important distinction between oral and literal domains, which must be better specified--at least into the domains of listening, speaking, reading/interpretation, and writing--for effective progress to follow. Moreover, each of these domains may require a distinct way of testing and deriving standards.

---

[1] National Educational Goals Panel, Report by The Task Force on Assessing the National Goal Relating to Postsecondary Education, July 31, 1992.

[2] For more information on the first workshop see, *National Assessment of College Student Learning: Issues and Concerns* and *National Assessment of College Student Learning: Getting Started*. Both are available from the Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328.

○ Improving postsecondary instruction is not merely another aspect of the situation that must be "factored in," but rather should be a major premise from which any assessment design and recommendations proceed.

○ Despite an ongoing and energetic debate about the context in which critical thinking skills can be discerned, another premise for their development and instruction is that universities organize their curricula around the distinction between general and subject-specific education, that both are essential to provide a foundation, and that higher level thinking must import a priori the assumption of competence in specific realms of knowledge and ways of thinking. Nonetheless, perhaps the most salient proof of critical thinking is the ability to transfer its process, as opposed to its products, to other realms of knowledge.

The 1992 Workshop: Preparation

Participants in the workshop that followed a year later in November of 1992 included the 4 authors of the primary papers, 12 reviewers, and about 40 people chosen to represent a national constituency of educators, professionals, policymakers, and a group of Washington-based professionals. The format of the workshop included both small groups and large plenary sessions. Four work groups were formed, each co-chaired by an author and a facilitator. Two of these groups concentrated on the problem-solving and critical-thinking skills, the third group on reading and writing skills, and the fourth group on speaking and listening skills. Each work group consisted of approximately 14 participants--an author, a facilitator, 3 reviewers for that particular author, 8 participants, and a recorder.

The categories of consideration for each group were far from arbitrary. The authors commissioned to develop recommendations through academic position papers were not only expert each in their chosen area, but were asked to establish their recommendations in the context of a broader view of that area. Since this type of groundwork had proven so useful with the first workshop, the 1992 effort was also founded upon this set of scholarly background papers, commissioned by NCES to describe specific aspects of the overall critical thinking assessment terrain.

It is important to note that the assessment of college student learning, as framed under Goal 5, Objective 5, is in the context of the global economy and in the practice of citizenship. Representatives from the business sector were in attendance at the first workshop. To assure that the skills needed for the practice of citizenship were considered, a number of scholars concerned with citizenship studies were invited to attend the second workshop, in addition to representatives from the business community. Each of the work groups had at least one person from this community of scholars. In addition, Susan Morse was invited to prepare a paper for the workshop, which she titled *Skills for Citizenship*. In the paper, she suggested that "many have assumed that being well schooled academically transfers to civic preparedness." This, she suggests, is not necessarily true. She warns, "As the needed competencies are identified for Goal 5, it is important to be specific about what we mean by civic knowledge and skills." She closes the paper with the challenge from John Dewey: "Democracy has to be born anew every generation, and education is the midwife."

As suggested above, the considerable work accomplished at the 1991 workshop had paved the way, and the categories established reflect the collective wisdom of the previous discussions about two major de facto features of the American educational scene: (1) no separate consideration was made of problem solving *per se*, for while academic distinctions of the study of problem solving persist, in the postsecondary world problem solving is the major *context* in which critical thinking skills are applied; (2) the phrase "effective communication" (inherited from the original goals) masks an important distinction between oral and literal domains, which must be better specified for effective progress to follow.

The four working groups were thus established as follows: two separate groups on problem solving and critical thinking, one on speaking and listening, and one on reading and writing. Another lesson learned from the 1991 workshop was that citizenship issues cut across all domains and can get lost when not specifically considered; thus a fifth paper was commissioned, identifying the skills needed for effective citizenship.

The authors had been asked to provide, for consideration by the workshop, their recommendations about three aspects in their particular area: (1) a preliminary listing of selected skills and the rationale for these choices; (2) levels of achievement or proficiency for each of the skills in their list; and (3) an approach to assessment of these particular skills--which might go so far as a suggested format.

Further, they were asked to consider the framework which produced Goal 5.5, namely that achievement levels should encompass not merely competence but rather accomplishment or mastery of the target skills, and should also cut across the three major contexts where improvement will be experienced: citizenship, suitability for the workplace, and graduate study. To take advantage of the important work already accomplished, NCES urged the authors to consult as source material the papers prepared for the 1991 Workshop, current literature and research, program and course syllabi, and institutional and state mission and goals statements. One further lesson learned from the first conference provided the authors an additional premise for their recommendations: that all college graduates will have had an opportunity to complete a general education core, in addition to a specialized or major area of study.

A methodology that proved successful the first year was revived: a second group of professionals who had strong ties to the specific content areas reviewed and commented on the four papers. Among the commentators chosen to provide reviews were job analysts, employers, baccalaureate and graduate school administrators and faculty, policymakers, researchers, and other communication and higher order thinking skills experts. Reviewers of the papers were asked to address the assessment of the higher order thinking and communication skills developed in the authors' papers, consistent with Goal 5.5, from the viewpoint of producing effective citizens and workers. Each reviewer commented on two papers--either both problem-solving/critical-thinking papers or the speaking/listening and the reading/writing papers. NCES asked the reviewers to comment on the comprehensiveness of the skills identified, the practicality of using the identified levels of proficiency for assessing the attainment of these skills, the recommended approaches for assessment, and the implications of using the suggested skills and levels of ability for the design of an assessment instrument.

Thus conference participants would arrive at the Washington workshop having been furnished a set of proposals, each supported by a scholarly treatise, in-depth reactions to these by selected reviewers, and the time to study and think about the implications and relevance of these ideas to their own experience and work. This effort to provide material in advance serves not only to focus concerns, but sets the tone for a true *working* meeting, designed to produce tangible results.

The Workshop in Action

On November 17, 1992, close to 70 people convened in the L'Enfant Plaza Hotel in Washington D.C. to consider and build on the material they had been sent: the original proposals of the five authors and the comments of the reviewers. As noted earlier, workshop participants were distributed among four working groups, each to focus upon one of the four scholarly reports (excepting the Morse paper on citizenship). Nearly 10 hours of individual discussion for each group were scheduled, in addition to several meals and meetings when everyone convened together. While each group was to develop its particular way of approaching the task, group members were initially presented by the workshop organizers with a fairly structured way to proceed.

NCES planners developed "The Icon Chart," essentially a device designed to solicit and systematize answers to the questions the group was to consider. As were the original authors, each group was charged with the task of producing a list of skills they believed 4-year college graduates should possess in order to meet Goal 5.5. The icon chart listed a sequence of questions, with the intention of leading the discussion towards an identification of domains and the macro/micro skills that serve them, and into an evaluation of how important and how feasible it would be to assess such skills. The four groups found the icon useful to varying degrees.

Each group met for several sessions over the 2 days, in the effort to reach a consensus on a listing of skills and related levels of proficiency for their particular area. The workshop's concluding session began with each of the authors reporting their reactions to the group's consideration of their paper. The floor was then available to all workshop participants to present, for the record, their views on their own working group meetings or on any of the basic questions originally posed to the authors. Finally, all participants were given the opportunity to comment on four questions, thus generating a survey of opinion about the workshop and the overall NCES effort.

Space limitations did not permit inclusion of reviewers' written comments in this report. However, in a number of instances, the remarks were included in the workshop reports or were addressed in the four papers, which were revised following the workshop.

Icon Chart of Higher Order Thinking and Communication Skills

Area Critical Thinking

Page____

A. Meanings and Unique Aspects of this Area	B. Dimensions, Categories, and Domains	C. Macro Skill Areas and/or competencies	D. Traits and/or Skills	E. C/W ?	F. Importance (1-10)	G. Feasibility (1-5)	H. H/L ?	I. Caveats and/or Concerns	J. Notes
1. CT/PS is purposeful and self-regulatory judgment reached by exercising generic cognitive skills in analysis, interpretation, inference, evaluation, and explanation.	1. ABILITIES 2. Cognitive	1. Verbal Reasoning skills 2. Argument Analysis Skills 3. Skills in Thinking as Hypothesis Testing 4. Using Likelihood and Uncertainty 5. Decision Making and Problem Solving Skills							
2. CT/PS is not speculation, meditation, contemplation, or appreciation.		6. Adaptability Skills 7. Creative Process Skills 8. Practical Procesing Skill 9. Wisdom							
		10. Inference 11. Evaluation 12. Analysis							
	3. Behavioral	13. Interpersonal							
	4. INCLINATIONS	14. Systematic Diligence 15. Intellectual Curiosity 16. Open-minded 17. Analyticity 18. Objectivity 19. Self-confidence							
	5. SENSITIVITIES								

CONTENTS OF THIS PAPER

This document portrays the essence of the foregoing events. **Part 1** includes **the skills papers**, the original material upon which the working groups built their discussions. These papers appear here not in their original draft, but rather as revised by each of the five authors, subsequent to the conference. This version thus reflects their thinking after considering the review comments and the discussions in their working group.

The working groups themselves were charged not with simply ratifying the list and schema proposed by their authors, but rather with using those as a point of departure to develop a product of their own. As suggested above, the icon as a useful device was perceived differently from one group to the next. The **group summary reports** included in **Part 2** were prepared after the conference, using notes and material developed in each group. They are intended to be fair and representative, though of course neither exhaustive nor definitive. The summary reports, to the extent possible, reflect the goal given to the group: to try to develop a list of skills and the ability levels that should accompany them. Additional material in this section:

○ Where **reviewers' comments** were germane to this orientation, they are included, a prime example of which are the remarks of Joann Carter-Wells on reading skills in the Group IV report.

○ The summary for working group III on **speaking and listening skills** has been supplemented by a listing of skills compiled in a special post-workshop session. Participants included the prime author John Daly; reviewer Andrew Wolvin; and Barbara Lieb, the facilitator for Work Group III.

○ Since **the Penn State delphi activity** was of prime concern and often cited during the workshop summaries, readers are referred to the comments of James Ratcliff, Director of Penn State's National Center for Postsecondary Teaching, Learning, and Assessment on plans for that activity in the proceedings for Work Group I.

Part 3 begins the process of evaluating the conference products. First, all participants were given the chance to take the floor and provide **comments for the record**. Next, a four-question survey was given to all and the **results of this survey** are here summarized. Finally, some **next steps** have been clarified and are outlined.

KEYNOTE REMARKS

by GARY PHILLIPS
Associate Commissioner for Education Assessment, NCES

This workshop is a part of an ongoing series of activities, steps that are underway at NCES, to implement eventually an assessment system, possibly a survey system, intended to measure critical thinking, problem solving, and communication skills at the college level. And this activity was actually on the books long before the goals were created at the meeting the governors had in Charlottesville in 1989, but it turned out that our thinking on this and their thinking were very much the same. And now, of course, the fact that we do have the six national goals has increased our awareness for doing this and also given us support for this activity.

If you'll recall, the goal we're referring to is Goal Five, which says that "Every adult American will be literate and will posses the knowledge and skills necessary to compete in a global economy, and exercise the rights and responsibilities of citizenship." And one of the Objectives of that Goal states that "The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially." And that's where we get the three domain areas of critical thinking, communication, and problem solving.

It also turns out that when the Goals panel formed their resource groups, the Interim Report on Goal Five stated that "If the National Goals Panel wishes to assess the ability of college graduates to think critically, to communicate effectively, and to solve problems, a new kind of assessment will have to be created. That assessment might be a type like the National Assessment of Educational Progress (NAEP) at the college level, given to a national sample of college students, at different kinds of institutions across the nation. To have credibility, such an assessment would have to take into account the differences in the postsecondary institutions in America, and the fact that the pluralistic system in place today has extended postsecondary education opportunities to the broadest cross-section ever of America's citizens. Developing such an assessment would be controversial for a whole host of reasons, and it could require as much as five years to develop, and cost several scores of millions of dollars."

They're right. It is controversial, and it will take a long time to develop, and we're just now in the initial steps. This meeting today is still a very initial, tentative planning session for this whole activity. The first step that we took was to hold a workshop like this, though broader in scope, last year in November. Over 70 people participated, we had a number of position papers, reviewers, practitioners, policymakers, and others in the field; you'll hear more about that later. One of the things that the workshop participants came away with is that the next step we should be focusing on is to try to hone in on the question of what skills we're attempting to measure, and to work on issues related to how to measure them. What the format should be, this sort of thing. We at NCES have accepted this approach, and our next step is to focus on a list of skills, and this is one of the foci of your meeting.

We believe at NCES that this assessment should have strong face validity: it has to make sense. We want it to be efficient. We want to have strong faculty support for it. We want to develop it so that it will be desirable for teachers to teach to it. And we want to have it consistent with the skills that students need, which we as a society recognize students should have. And of course we also want to be sensitive to the diversity found in the postsecondary environment. The next step, after this meeting, is that the project goes on for further development work with a project at Penn State, which will be discussed later.

One of the commissioned papers for this conference, by Steve Witte, expressed the fear that the search for a national consensus may create a camel, when what we want is a horse. I have a fear that if we listen to a narrow constituency we could end up with a turkey. I think it is possible to obtain a national consensus and still get a horse. We have done this for 20 years with the National Assessment. It is still thriving, in fact I think it's a thoroughbred. So I think this can be done, although it will take a lot of hard work, a lot of effort, a lot of money, and a lot of time.

I don't think any of us are intending to rush into this. We do want to take the time and the effort and to put in the hard work that it's going to require. I want to introduce Emerson Elliott, the commissioner of NCES. Emerson was serving as acting commissioner for a while, but we're all glad he's now for real. He's back from a speaking engagement at Harvard, looking very much an ivy-leaguer, and so Emerson, I'll turn it over to you.

by EMERSON ELLIOTT
Commissioner of Education Statistics, NCES

Thank you, Gary. Well Gary has I think very well set the context for this meeting, and Jeff will tell you the details of what we expect to accomplish, so I don't know that I have a lot to talk about. But let me make some comments, some of which a few of you have heard me make before, and some of these are additional thoughts about the context. I think your time here is very highly structured, and I think that was exactly the intent. As we face this important task of working through a set of issues to help us figure out how we might measure skills does require a highly structured meeting.

We're talking about a college NAEP, and I guess that's shorthand. Whether it will end up looking like NAEP is not necessarily all that clear, but I do want to tell you from my day in Cambridge that research in higher education, at least at Harvard, is alive and well. And let me give you an example I picked up from Professor Richard Light, a professor of statistics at the Harvard Graduate School of Education. He studies teaching and instruction at Harvard University.

He was telling about his lecture in statistics, which was reaching what he described as "the most difficult concept of all in statistics" that he wanted to teach. It was on sampling, but I don't remember which aspect of sampling it was. He had arranged for a friend of his to come to this class, someone very adept in precisely this element of statistics. He wanted the guy to come to the class, sit in the back of the room, and read a newspaper, riffling the pages noisily while Dick was giving his lecture on this most difficult of all statistical concepts. He told me that you can't tell who is a student here because there are people our age who are students, coming back to Harvard to work on a doctorate after they're well launched on their careers, so nobody could pick up anything amiss from the age of Dick's paper-riffling friend.

So this guy sat in the back of the room, rustling his paper, and flicking it from time to time to make more noise. And Dick went through his entire lecture, and had everyone's attention, having warned them the subject matter was so very difficult. The students, of course, were noticing all of this rustling in the back, and finally, at the end of the hour, Dick said, "You, back there. Hey, you. You with the paper, did you understand any of this? What I've been saying here?" "Well, yeah," came the answer. "Well why don't you explain it to the people here?" he said and then had this guy get up and give a 4-minute summary, which he said was the clearest 4-minute summary imaginable, which demonstrated elegant thinking on this particular concept. And he concluded the class without ever telling anyone this was all rigged ahead of time. So that's the way they do it at Harvard.

Now I'm not sure quite what that says about doing assessment in higher education, but maybe there's something to learn. What we're all about here is really very challenging to the measurement community, and also very challenging to higher education. Gary has briefly described several principles that we're trying to achieve. And you know we stated those following the meeting last year, so for those of you who were here at the first conference looking at overall issues and concerns related to Goal Five, Objective Five, we listened. If we are saying what you think you meant to say, that's fine. If not, you'd better let us know.

There are principles that we think are very important to incorporate into some eventual assessment system in higher education. One is that we're determined to design it in an open forum. So that means we will continue to have conversations of this kind. We're very pleased that we were able to make things work out with the National Center for Postsecondary Teaching, Learning, and Assessment to work on a follow-up that will involve many more people in a Delphi technique and continue to pursue these issues. But it has to be done in an open and participatory way.

And the participation has to involve faculty. That's not something that NAEP necessarily set out to do, and that's why I'm trying to draw some distinctions here. One of the principles is --since it's been 30 years since NAEP was conceived--that perhaps we have learned a few things in the meantime. The political debate has changed. The kinds of people who are interested in this kind of measurement are now different from those who were interested when the National Assessment was first created, and the state of the art is different. So a strong faculty participation is essential because, if in fact faculty don't believe that the kinds of measures we finally prepare are important, it simply will not work. At least in my mind it won't.

Along with that, the political leadership also has to agree that this is important and worth doing. First of all, they won't provide the money if they don't. But second, if there isn't a general consensus, if they don't come to understand why the faculty come to look at things the way they do, and why other people who look at measurement want to do it the way they have proposed, they won't support it. And they will continue to press for something that is probably multiple-choice and simple and fast and machine-scorable, and we all know what we think about that.

Now Gary slid very quickly over "teaching to the test" being desirable, but that's code for saying that the intent is that the tasks finally emerging from this should be tasks people believe are worth doing, and worth teaching. And that in fact they are valid measures of what it is that we want students to learn. And if we're not able to say that, then a great deal of the value of this task will have been taken away. We also want to minimize the number of test items, and that's one of the places where I think measurement people will have to work very, very hard, and it's one of the things that makes it especially challenging to them. Because we do want to reduce the overall burden from this. On the other hand, we want to maximize the amount of learning from the data. So I think the tradeoff is in the direction of a smaller number of items that will tell us more, but perhaps at the expense of covering everything that we would like to have people know. That is, we will have to come to be more specific about that portion of the curriculum that we're testing.

Now if NCES fails to follow these principles, I do hope that you will let us know, in this conversation and the ones that come later. And if you think these principles are wrong-headed, I hope you also will tell us, that is, whether they are incorrectly formulated. Let me just conclude by saying that the challenge to the measuring folks is whether it is possible to have something that is both reliable and valid, and meets all of the criteria that I have just described. And that is no easy job. I think for higher education, it is equally challenging, because it demands that higher education make a statement about what they believe is important, which higher education has really not wanted to do. Higher education has generally wanted to say, "Well, in this institution, these things are important, but in another institution they're working for a different clientele, and they have a different set of objectives." And people just don't want to bring that together. So it's very challenging, this agenda, for both higher education and for measurement, but I know you're up to the challenge, and I know you came to this conference to help us move along on the challenge. And I wish you well.

SKILLS FOR CITIZENSHIP

by Suzanne W. Morse

National Education Goal Five mandates that Americans possess the knowledge and skills necessary to exercise the rights and responsibilities of citizenship. The idea that specific skills and knowledge are needed to practice democratic citizenship is not new. In fact, it was a point debated by the Founders. At issue early on was whether the mass citizenry had the capacity to be directly involved--to practice citizenship. Thomas Jefferson was the person who felt most strongly that citizens could and should be involved in governing.

Jefferson was quite clear on the issue.

> *I know no safe depository of the ultimate powers of the society but the people themselves; and if we think them not enlightened enough to exercise their control with a wholesome discretion, the remedy is not to take it from them, but to inform their discretion by education (1903, vol. 8, p. 278).*

This statement carries two messages: the first is directed to education itself, saying that people need general knowledge to participate; the second refers to the necessity of specific skills in order to fulfill the requirements of citizenship. For the purposes of Goal Five, the messages are not mutually exclusive. Citizens must be able to understand the issues, decide together appropriate courses of action, and, finally, take action.

The rights and responsibilities of citizenship are not automatically known in democratic societies--they are not learned by osmosis. While place of birth gives one citizenship, it does not make one a "citizen." That distinction puts people on notice that while they hold title to residency by virtue of the place of their birth or their parents' birth, they must know and do something specific if they are to be practicing citizens. There are requirements and responsibilities. The way people exercise their citizenship is predicated on how they understand these requirements and responsibilities. That is, if a person considers informed voting as the sole requirement and responsibility of citizenship, then his or her preparation and subsequent actions will follow a certain path. Or if people see only the entitlements of citizenship, then their actions and expectations may take a nonparticipatory course. Jefferson, however, had a broader concept of democratic citizenship--one that entailed both representation and participation. That vision requires specific preparation and action. Developing a personal concept of citizenship in a democracy is part of the educational process itself. It is critical that if citizenship is to be renewed, an awakening of the "office of citizenship" must occur (Ketcham 1987, p. 145).

Informing Their Discretion

Education for citizenship has come to be synonymous with a general education. Many have assumed that being well schooled academically transfers to civic preparedness. It is true that abilities such as deliberation and judgment can be transferred, but their political applications have different purposes and different intents. As the needed competencies are identified for Goal Five, it is important to be specific about what we mean by civic knowledge and skills.

Skills for citizenship are distinguished from other skills because we are free people who live in a society that is governed by the people. John Adams, though conservative in other views, gave real insight to the necessary civic skills when he said "Liberty according to my metaphysics is an intellectual quality...it is a self-determining power in an intellectual agent. It implies thought and choice and power" (cited in Greenstone 1991, p. 33).

Adams' idea of the practice of freedom supersedes any prescribed pedagogy. Rather it points the way to an intellectual activity that includes the positions and thoughts of others. This is in harmony with what we think of as basic preparation for citizenship. Education for citizenship must go beyond courses to conceptual skills that are needed to grapple with and function in the modern world. This is consistent with what we hope happens in the liberal arts. As Thomas Jeavons wrote in *Learning for the Common Good* (1991), at the heart of the goals for liberal education is the development of the capacity for understanding and sound judgment.

However, it is the distinction between public skills and personal skills that demands attention. It is one thing to have a vivid personal imagination, it is quite another to have public imagination. Making that distinction does not require a different body of knowledge or separate courses, rather it commands a different perspective that can be brought to every course and experience. It is the difference between asking "What should I do?" and "What should we do?"

One of the primary issues associated with civic education is its relationship to the overall curriculum. In order to accomplish the educational goals mandated, it is essential that civic education not be a shadow or parallel curriculum or a collection of courses. Rather it must be thought about as pervading every course and discipline and the academic community must think about it consciously and deliberately. As Derek Bok wrote, "We ought not mystify the notion of civic education to the point that it can be done only by special people in special ways under special circumstances" (1985, p. 48).

What Citizens Need To Know

The early examples of democracy, however flawed, provide a legacy for the relationship of people to their community and a blueprint for what citizens should do. Although ancient Athens was "the most contradictory, the most paradoxical society imaginable" (Sagan 1991, p.2), it emphasized the particular relationship of citizens to the community—so much in fact, that those who did not participate were called idiotes.

Therefore, to understand what citizens should learn, we must first determine what citizens in American democracy should be able to do. First, they must be able to decide among a range of issues that relate to public policy. There are those who believe that these decisions should be left to elected representatives. Even so, citizens cannot abdicate their ability to consider options and weigh alternatives. Second, citizens must have the capacity to know and work with others on areas of common concern. This means knowing about issues, but also knowing who in the community or organization should be at the table before a decision is made. Citizenship or public action does not imply friendship or collegiality; rather, it is often a joint effort among strangers.

The basic skills for political life require a perspective as well as a mastery. Students must understand that the work of democracy is messy, inefficient, and frustrating. These very traits separate the American system of government from the early days of monarchical rule. Unfortunately, over the last several decades, the country has given way to mass representation in which most citizens are excluded from any level of direct decision making, a situation exacerbated by the complexity of the issues, the actions of special interests, and the powerful influence of the media. This disconnection from public life causes individuals to isolate themselves more and more and pursue their private interests. Two recent accounts confirm this idea. E.J. Dionne Jr.'s book, *Why Americans Hate Politics* (1991), and the Harwood Group's report, *Citizens and Politics: A View From Main Street America* (1991), both made it clear that citizens are alienated by politics today. Dionne wrote that politics defines issues in such partisan ways that finding agreement is nearly impossible. Citizens are polarized from the start by the way the choices are presented. Americans feel, Harwood reported, not apathetic but impotent--pushed out of the political process. His study found that Americans care deeply about public life but do not know where or how they can participate. Americans are looking for ways to participate. These themes resonate with younger Americans as well. An article in *The Washington Post* painted a gloomy picture:

> *...what defines members of the age group (18 to 24) even more is that in a year of political alienation, they are more pessimistic about the future and disengaged from politics than any other group (Morin and Balz June 17, 1992, p. A16).*

Given the challenges of the current political scenario in America, it is important that citizens and particularly younger citizens know what is expected of them in a democracy and how those expectations affect what they do. Mr. Jefferson warned of a situation where free citizens had no outlet or opportunity to exercise their civic skills.

> *Jefferson knew, however dimly, that the Revolution, while it had given freedom to the people, had failed to provide a space where freedom could be exercised. Only the representatives of the people, not the people themselves, had an opportunity to engage in*

> *those activities of "expressing, discussion, and deciding" [that] in a positive sense are the activities of freedom (Arendt 1965, p. 235).*

What then does citizenship require? Briefly, five key skills are at the root of practicing citizenship: **political talk, political thinking, political judgment, political imagination, and courage to act.**

Political talk is not just speech or talk--it includes a whole range of communication skills: listening, cognition, setting agendas, and mutual inquiry. According to Benjamin Barber in *Strong Democracy*, "Political talk is not talk about the world. It is talk that makes and remakes the world" (1984, p. 177). Said simply, talk is the necessary foundation to build a community. Good civic conversations:

> *....start with the assumption that, while we all have opinions and can have pertinent facts, good talk does not consist of a trading of, or negotiating between, such pre-existing stances and evidence...At its best, the public is where we find, through conversation with others, a deeper level of understanding of our opinions and a level of understanding of ourselves and others (Minnich 1988, pp. 34-35).*

To learn this basic skill of democracy, citizens must talk with each other. Public talk itself becomes the pedagogy. Pericles had this conception of talk when he said that the Athenians "taught themselves first through talk" before taking action. Talk is the linchpin in civic life that allows citizens to define, create, dissent, confirm, and act. Thomas Jefferson writing at age 70 to John Adams, age 78, expressed the essence of talk in a democratic society:

> *I state my difference with you not because I wish to begin a controversy when we are both too old to change opinions arrived at over a long life of experience and reflection. I believe that we ought before we die to explain ourselves to one another (Hutchens 1982).*

Political thinking is not a singular activity but a collective skill requiring mental interaction with others. Developing the ability to think politically is required for citizenship. Taken from the study of critical thinking, the process includes three primary elements: (1) an attitude to consider thoughtfully problems and issues; (2) knowledge of logical and reasoning inquiry; and (3) the ability to think with those not present (Paul 1984, pp. 13-14). Essential political thinking not only makes citizens consider issues in more inclusive ways, but also connects them to the larger world. Arendt called this skill the "enlarged mentality." We must think in the place of everybody else.

While critical thinking and political thinking are not synonymous, they do share many of the same elements. Political thinking, however, expands to include the perspectives of others--what Arendt might call the "thinking with others" capacity.

The skill of thinking in a healthy democracy, in addition to helping individuals consider specific events in life in a more reasoned way, has the psychological dimension of making citizens feel connected to the larger world. The connection of critical thinking to an individual's civic life is one of the "necessary tools and skills [students need] to make sense of the social relations, material conditions, and cultural milieus in which they exist and their relationship to wider society and dominant rationality" (Kretoric, cited in Brookfield 1987, pp. 53-54).

Political thinking is the basis for our capacity to judge, imagine, and act in a public way. The ability to think critically, and one could add publicly, is recognized by "a readiness to ask why things are the way they are, a capacity to speculate imaginatively on alternative possibilities, an inbuilt skepticism of the pronouncements and actions of those who are judged to be in positions of political and economic power..." (Brookfield 1987, p. 68).

Talking and thinking are building blocks for another skill required of participatory democracy--**political judgment**. Political or public judgment is not personal judgment. Individual judgment is concerned with what "I" will do: public judgment is about what "we" will do. Public judgment is first and foremost a process to deal with the inevitable uncertainty in public life; it is required to sort through the ambiguity created between opinion and judgment. Hannah Arendt gave a practical example of this concept's application in the modern world in her work on the war crimes trial of Adolph Eichmann. She said of Eichmann that he lacked reflective judgment. For even in his trial, Eichmann said that he had done nothing wrong in overseeing the massacre of hundreds of thousands of Jews. He had no capacity to see beyond following orders. He said he would have felt guilty otherwise. The process of judging requires that others be recognized and considered in the process of thought and action. Judgment is "the capacity to accept the human condition of plurality...Judgment (or taste) decides not only how the world will look, but who belongs together in it" (Arendt 1968, p. 223). Public judgment is required because politics is about the uncertainties that surround public issues. Rarely are there clear right or wrong responses. Rather citizens and policymakers are faced with the fundamental task of making a choice between two or more options. The ability to choose and solve problems requires the faculty of judgment. In the ultimate analysis, there is expert testimony for all sides of all issues. The citizen is left to balance the common interest against the choices available. Because public judgment implies that what we can know together as citizens we cannot know alone, it is imperative that students be exposed to diversity, pluralism, and notions of a common agenda. We study judgment in order to identify and strengthen a mental faculty that allows citizens to go beyond rules, regulations, and technical expertise to sort out issues for themselves. Judgment allows the "opening of a space of deliberation that is being closed ever more tightly in democratic societies" (Beiner 1983, pp. xi-xii).

The skills of thinking and judgment are joined by the ability to **imagine**. Imagination is "that strange ability to make present what is absent and to make ourselves absent from our immediate perspective and present to some abstract perspective. Through this process, we are able to put ourselves in the other's position and see, not as he sees, but how it looks to us from his point of view" (Denneny 1979, p. 264). Imagination makes judgment possible.

In relation to civic participation, citizens need the ability to imagine how the world could be different. "Judging...activates imagination by demanding that participants reexamine their values and interests in light of all the inescapable others--the public" (Barber 1984, pp. 136-37).

Advocating imagination's role in the civic development of students assumes that a primary goal of education is to develop the capacity of judgment so that students might imagine a different world (Smith 1987). Exposure to different perspectives, different points of view, and different ideas stimulates students to think of worlds and societies not yet visited or created. As Kant argued, the imagination is the condition of all thought and makes all other cognitive faculties possible (Kant 1951).

Beyond all these skills for a participatory democracy--talk, thinking, judgment, imagination--is the **courage to act.** Because of the uncertainty in the political realm, action based on judgment requires the courage to act. Arendt said that courage is the political virtue because other skills become worthless without it (Denneny 1979). Courage is best explained through example. Exemplary courage was displayed by the Danish citizens during World War II. During the early German occupation of Denmark, there had been no attempt to deport Jews to concentration camps as in other parts of Europe. However, in August 1943, the tension had reached an all-time high in the country with the resignation of King Christian X and numerous acts of citizen rebellion on the part of the Danes. On September 28, 1943, an anti-Nazi, German embassy official, G.F. Duckwitz, secretly sent word that on October 1-2--the beginning of the Jewish holy days--the Jews would come under attack. Within 48 hours, more than 7,200 Jews had been contacted, hidden, and transported to safety by their non-Jewish fellow citizens. Only 464 Danish Jews were taken to the Theresienstadt concentration camp. This remarkable story illustrates both judgment on the part of the Danes and the courage to act on that judgment (Merton 1971, pp. 166-67).

> *The Danes were able to do what they did because they were able to make decisions based on clear convictions about which they all agreed. Action became possible when fundamental truths were taken seriously (Merton 1971, p. 167).*

How people have the courage to act has everything to do with how they think of politics and their role in it. The courage to act comes from the interaction of the other civic skills. In this nation, the Civil Rights Movement was predicated on the civic capacity of a group of people to think, judge, imagine, and then act on a societal problem. There were events and stimuli that caused certain actions, but the movement itself came from long years of practicing civic skills in hiding. The Southern Christian Leadership Conference's program on citizenship education and programs at the Highlander Center in Tennessee and elsewhere were instrumental in helping to develop those skills and taking them from private to public.

Outstanding individuals like Martin Luther King, Jr., Rosa Parks, Fannie Lou Hammer, and Ralph Abernathy were critical, but the movement was a collective citizen effort (Morse 1989, p. 77).

The courage to act comes from several sources. It can come from the student's own sense of right and wrong; it can come because of the example set by others and with others; and finally, it can come because of tested cause and effect experiences. As the Harwood study found, people act because they feel as if their actions will mean something or will alter the way things are done. Therefore, it is imperative that students gain experience talking and thinking with others in order to solve a public problem. Where the experiential movement often falls short in preparing people for citizenship is when the experience does not take into account the public dimension. While the experience may inform and inflame, it may not teach public skills.

However, there is no one way to learn and teach about politics and the citizen's role in the political, short of practice. But practice does not mean just doing anything; rather, it means doing activities based on an agreement of how a democracy or a community should function. Living in concert with others is the essence of a democratic community. This is where the connection between education and democracy begins:

> *The point where democracy and education intersect is the point we call community. For if democracy is a mode of associated living, then it is also true, Dewey has written, that 'in the first place, the school must itself be a community life!' (Barber 1989, p. 69).*

The pedagogy for civic education must include theory and practice. It includes both what is taught and how it is taught.

Role of Higher Education

The preparation of the next generation of citizens has received considerable attention in recent years. There is the paradoxical situation of low voter turnout in the 18-25 age group juxtaposed against a burgeoning community service movement. What this paradox shows is not unlike what we have seen in other generations--a mix, albeit unhealthy, of the participators and the spectators. What makes the current situation more grave, however, is that today's students are facing a world with enormous problems and demands.

The twenty-first century presents challenges different from any we have seen in modern times. Those whom we thought were enemies may not be. The unlimited resources that we thought would outlast us may not. The need for services far outweighs the capacity to respond even to the youngest and the frailest.

While higher education cannot be expected to solve our public ills or produce all the people who will, it has the unique opportunity to provide the theory, the practice, and the place for civic learning. The college years allow students the freedom to develop the civic skills they will need throughout life. Colleges and universities are civic communities. They are assemblages of diverse people living together who are confronting the same kinds of issues as people outside the campus. In fact, higher education communities have become the focal points in recent years of some of the worst aspects of public life: racism, sexism, violence, and a pervading sense of powerlessness. This very fact gives them particular currency as laboratories for students to learn and practice an expanded notion of citizenship and of themselves.

However, for the process of preparing students for citizenship to be effective, colleges and universities must decide together the route to take. All too often, the task of citizenship gets passed to one department or office. This is a responsibility for the institution as a whole and in order to make the decisions and chart the course, a civic conversation must take place. This is the legacy of higher education. John Dewey summarized well the challenge:

> *Democracy has to be born anew every generation, and education is the midwife" (cited in Curti 1965, p. 499).*

References

Arendt, Hannah. 1963. *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Viking Press.

______. 1965. *On Revolution*. New York: Penguin Press.

______. 1968. *Between Past and Future: Eight Exercises in Political Thought*. New York: Penguin Press.

Barber, Benjamin R. 1984. *Strong Democracy: Participatory Politics for a New Age*. Berkeley and Los Angeles: University of California Press.

______. Fall 1989. "The Civic Mission of the University." *Kettering Review*: 67-72.

Beiner, Ronald. 1983. *Political Judgment*. Chicago: University of Chicago Press.

Bok, Derek. Winter 1985. "The University and the Publics Beyond its Walls."

Interview by Bernard Murchland. *Kettering Review*: 46-52.

Brookfield, Stephen D. 1987. *Developing Critical Thinkers: Challenging Adults to Explore Alternative Ways of Thinking and Acting*. San Francisco: Jossey-Bass.

Curti, M. 1965. *The Social Ideas of American Educators*. Paterson, NJ: Littlefield, Adams & Co.

Denneny, Michael. 1979. "The Privilege of Ourselves: Hannah Arendt on Judgment." In *Hannah Arendt: The Recovery of the Public World*, edited by Melvin A. Hill. New York: St. Martin's Press.

Dionne, E.J., Jr. 1991. *Why Americans Hate Politics*. New York: Simon & Schuster.

Greenstone, J. David. 1991. "Adams and Jefferson on Slavery: Two Liberalisms and the Roots of Civic Ambivalence." In *The Constitution of the People*, edited by Robert J. Calvert. Lawrence: University of Kansas Press, p. 33.

Harwood Group. 1991. *Citizens and Politics: A View from Main Street America*. Dayton, OH: Kettering Foundation.

Hutchens, Robert. 1982. "The Aims of a Democratic Society." *Center Magazine*, (March/April), back cover.

Jeavons, Thomas. 1991. *Learning for the Common Good.* Washington, DC: Association of American Colleges.

Jefferson, Thomas. 1903. *The Works of Thomas Jefferson*, edited by Paul Leichester Ford. 12 vols. New York: Knickerbocker Press.

Kant, Immanuel. 1951. *Critique of Judgment*, translated by J.H. Bernard. New York: Hafner Publishing.

Ketcham, Ralph. 1987. *Individualism and Public Life.* New York: Basil Blackwell, Inc.

Merton, Thomas. 1971. *The Nonviolent Alternative.* New York: Farrar Straus Giroux.

Minnich, Elizabeth Kamarck. Summer 1988. "Some Reflections on Civic Education and the Curriculum." *Kettering Review.*

Morin, Richard and Dan Balz. 1992. "Children of the Tuned-In Find Politics a Turnoff." In *The Washington Post.* June 17, pp. A1, A16.

Morse, Suzanne W. 1989. *Renewing Civic Capacity: Preparing College Students for Service and Citizenship.* ASHE-ERIC Higher Education Report No. 8. Washington, DC: School of Education and Human Development, The George Washington University.

Paul, Richard. 1984. "Critical Thinking: Fundamental to Education for a Free Society." *Educational Leadership* 42(1): 4-14.

Sagan, Eli. 1991. *The Honey and the Hemlock.* New York: Basic Books.

Schlesinger, Arthur, M., Jr. 1986. *The Cycles of American History.* Boston: Houghton Mifflin.

Smith, Jonathan. November/December 1987. "Playful Acts of Imagination." *Liberal Education* 74: 14-20.

A NATIONAL ASSESSMENT OF CRITICAL THINKING SKILLS IN ADULTS: TAKING STEPS TOWARD THE GOAL

by Diane F. Halpern

In 1989, the National Education Goals Panel announced six goals for American education that, if achieved by the year 2000, would help us build a "nation of learners." The purpose of the National Education Goals Report was "to focus on measurement" (Romer 1991, p. iii) so that Americans can chart progress toward these critical goals. Only one of these goals, Goal 5, pertains to the education of adults. Goal 5 targets the development of the adult-level critical thinking skills that are needed to enhance and sustain a strong economy and the responsibilities of citizenship.

> Goal 5. By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

This proposal is in response to Objective 5 of this goal:

> The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially (National Education Goals Panel 1991, p. 237).

It is a proposal for a comprehensive plan to examine the critical thinking skills of college-educated American adults and to chart progress toward enhancing these skills. Without a program for assessing progress, the Education Goals are hollow rhetoric; without action to help Americans achieve these goals, they are empty promises.

Section I. The Case for a National Assessment of Critical Thinking Skills in Adults

What proportion of American adults have the knowledge and skills that are needed to enhance and sustain a strong global economy? How can we improve postsecondary teaching and learning so that more Americans obtain a high level of competency in the abilities to think, learn, and communicate? Do American adults (both college graduates and those without college degrees) compare favorably with adults in other countries in the knowledge, skills, and abilities that are likely to be needed in a rapidly changing world? Has the emphasis that U. S. colleges place on individual achievement adversely affected the willingness of college-educated adults to work cooperatively? How can we determine what works in higher education?

These are crucially important questions, and the way we answer them will have a profound effect on the type of nation that the United States will become in the rapidly approaching 21st century and beyond. These are questions about adult-level cognitive achievement and ways to promote cognitive development in college classes. Although Objective 5.5 specifically targets college graduates, it is necessary to collect information about adults who have not attended college and adults who have attended, but have not completed college, in order to provide appropriate comparisons. Thus, information about the critical thinking skills of adults who are not college graduates will also be examined in this proposal.

All of the questions posed require a thoughtful examination of what adults know and need to know so that Americans can compete and cooperate with citizens from other countries at a time in history when planet Earth is growing smaller and life on other planets is no longer in the realm of science fiction. These questions also share another common feature--a national assessment of the critical thinking skills of adults is the only way to answer them.

A national assessment of the critical thinking skills of adults offers many potential benefits, but it is important to remember that assessment, for assessment's sake, is an expensive and futile exercise. Like any other measurement, it is not an end in itself, but a means to an end. When assessment is done well, and the results are used to determine the educational practices that are linked to improvements in the number of students who think well, then it can provide valuable information for achieving and documenting the National Education Goals. Thus, the two essential components of a quality assessment are that it provide valid and useful data and that the data be used to establish goals, rank priorities, and direct educational reform.

Improving Teaching and Learning

There are many reasons for assessing the higher order cognitive skills of adults. Perhaps the most persuasive reason is that an assessment will provide feedback that can be used to inform and improve postsecondary education. America's colleges and universities have no objective way of determining if they are doing a good job of developing the thinking abilities of their students. A national assessment of critical thinking will fill this void by providing a yardstick with clearly marked competency levels that can be used to see how well their students "measure up." There can be no meaningful discussions of improving education at the college level without baseline data that tell us where we currently are (Nummedal 1991). Students also need these indicators so that they can identify their strengths and weaknesses and use this information about their thinking skills and abilities as a guide to career and course planning.

A quality assessment will also provide a fine-grained analysis that reflective institutions can use to identify the skills that they are successfully developing in their students and those skills that their students are not learning. This sort of skill-specific results will provide the information that is needed to close gaps and avoid redundancies in educational programs. Without information of this sort, there can be no data-based decisions to direct curricular or other educational reform.

A national assessment can enrich our understanding of the processes that underlie college-level teaching and learning. It will provide the data that cognitive theorists need to posit and revise models of adult cognitive development. Little is known about the way cognitive abilities change over the adult life span prior to old age. A large-scale carefully executed assessment program will provide a wealth of data that can direct theory and research about adult cognition well into the next century. The information that is collected can describe the outcomes of current educational practices. Educational theorists and researchers can use these data to explain how adult learning occurs and to identify ways to improve it.

A national assessment of critical thinking skills is a powerful tool that can be used to improve teaching and learning at the college level. Banta (1991, p. 29) enthusiastically described a high quality national assessment of critical thinking as "the boldest and potentially most promising research and development project ever undertaken in higher education."

The focus on critical thinking is a particularly important area of educational research because there are rapid changes in what adults need to know as the workplace undergoes radical technological and social changes. There continues to be a virtual explosion of information in every contemporary arena as modems can now deliver the contents of entire libraries to a worker's desk and other electronic information delivery systems zap information between distant sites at rates that are faster than the eye movements that take this information in. More than ever, adults must know how to analyze and evaluate large quantities of rapidly changing information and seek and generate novel solutions to problems that didn't exist when they were in school. Today's students may ultimately work at jobs that have no current counterpart. They must be taught both how to think clearly about complex issues and how to learn new skills and acquire new knowledge.

Accountability

The cost of higher education is expanding while, at the same time, funding has been sharply curtailed (U.S. Bureau of the Census 1991). Taxpayers, state legislators, employers, prospective students and their parents, and all of the other stakeholders in higher education have been demanding evidence that their scarce education dollars are being well spent. Gold, Director of the Center for the Study of the States, summed this situation up when he commented, "Education is going to have a lot more competition for dollars (quoted in Brownstein 1991, p. A5). A national assessment that examines gains in critical thinking skills that are attributable to the college experience can go a long way toward demonstrating that college produces tangible positive results in students' ability to think critically. A valid and reliable program of assessment can provide valuable information about the impact of college on cognitive development, thereby providing strong support for those who want to keep America's successful colleges fiscally sound.

The Crisis in Human Capital

The national data on the knowledge and skills of America's adolescents and young adults are grim. An abysmal conclusion about the scientific thinking skills, based on data from 1.4 million students, was recently published by the Educational Testing Service: "The majority of high school graduates are unable to understand 'relatively complicated' information, use basic algebra, or evaluate the appropriateness of scientific procedures" (De Loughry 1989, p. A2).

Although we have little data on the social consequences of the overall low level of academic achievement among young adults, the economic consequences of failing to produce citizens who are educated in scientific reasoning are alarming. For example, Walberg (1989) reports that there is a significant correlation between the science test scores of 14-year-olds and national economic growth a decade later. Other sources of data relate science learning, research and development activities, and economic indices such that all three are inextricably linked. The implication is clear: if our young people do not have an advanced level of the knowledge and skills that underlie research and development in technology, then our economy will suffer. It is essential for us to address the issue of the development of scientific and other types of critical thinking in order to maintain our position as a world leader in science and technology and to heal our ailing economy. Thus, another benefit of a national assessment of critical thinking is that it will provide numerous opportunities for greater collaboration between the inhabitants of two different "worlds"--the world of work and the world of higher education. Although these two worlds may be in close proximity in space and time, they rarely communicate.

American high school students routinely rank below those from other industrialized countries in academic areas such as scientific knowledge, mathematical problem solving, and general literacy (Alsalam, Ogle, Rogers, & Smith 1992; Lapointe, Mead, & Phillips 1989). We currently have no comparable knowledge about college students, college graduates, and other adults because adult learning and knowledge have been neglected components of earlier education initiatives. Economists, politicians, business leaders, and others are now concerned with the "crisis in human capital"--shortages in America's most precious commodity--educated, thinking adults. As the number of unskilled jobs shrinks, undereducated adults line up at employment offices while openings in engineering and other skilled positions go unfilled. Occupational forecasts show that "there is a steady increase in those occupations that currently employ more highly educated workers" (Bailey 1991, p.11). Although this point is addressed more specifically in the Education Goals Panel Goal 4, Objective 3, a national assessment of higher order cognitive skills will help us to achieve this portion of Goal 4 as well.

Creating Self-Fulfilling Prophecies

A final reason for embarking on a national assessment of critical thinking skills is that it sends a clear message to all of the participants in higher education that better thinking is an intended outcome of college attendance. When students, faculty, and administrators know that student thinking skills will be assessed, a renewed emphasis on the promotion of these skills will result. Students will know early in their college career that they are expected to improve in their ability to think clearly and that tangible, observable evidence of this improvement is an anticipated outcome. Faculty will discuss ways to enhance cognitive skills in their courses and administrators will fund projects that have this goal as their focus. In short, a national assessment of critical thinking skills sends a strong message to the higher education community that good thinking is valued, assessed, and rewarded. All of the players in higher education should act in overt and covert ways to bring about the desired change.

Section II. What Are Critical Thinking Skills?

Although there are many definitions of critical thinking, most of them include skills in applying, analyzing, synthesizing, and evaluating information as well as the disposition to use these skills (National Center for Excellence in Critical Thinking Instruction 1991). It is important to note that critical thinking comprises at least two parts--skills and the disposition to use these skills. A quality assessment will examine both of these components. A general list of critical thinking skills that would be applicable in a wide variety of work settings and in the exercise of citizenship would include: understanding how cause is determined, recognizing and criticizing assumptions, analyzing means-goals relationships, assessing degrees of likelihood and uncertainty, recognizing problems, incorporating isolated data into a wider framework, and using analogies as an aid to decision making. Specific lists of skills that have been categorized along several dimensions can be found in several sources (e.g., Facione 1991; Halpern 1989; Paul & Nosich 1991).

Rationale and Criteria for Targeted Skills

An assessment of critical thinking is predicated on two underlying assumptions: (1) that there are clearly identifiable and definable thinking skills which students can be taught to recognize and apply appropriately, and (2) if the skills are recognized and applied, the students will be more effective thinkers. This means that an outcomes assessment of critical thinking will need to have operational definitions of those skills that are being examined along with a clearly articulated rationale for the selection of those skills that are targeted for assessment. The following criteria are suggested for selecting skills for a national assessment of college-level critical thinking skills.

1. Workplace and Citizenship Skills

Goal 5 established the criterion that the skills that are to be assessed are those that are needed to compete in a global economy and in the exercise of citizenship. Workplace skills were identified in a report from the Secretary's Commission on Achieving Necessary Skills (1992). In this study of workplace skills and knowledge, thinking skills were defined as "the ability to learn, to remember, to reason, to think creatively, to make decisions, and to solve problems" (p. xiii). Thus, it seems reasonable to use this definition as a starting point for identifying critical thinking skills in the workplace and require that any skill that is targeted for assessment be applicable to one of these broad categories.

The thinking skills needed for the exercise of citizenship have also been the focus of recent reports. Morse, in her paper entitled "Skills for Citizenship" discussed those skills needed "to consider thoughtfully problems and issues" of citizenship. In another paper by Morse (1989), she calls critical thinking a prerequisite for civic responsibility. She lists the critical thinking skills that are necessary for high level citizenship as "an attitude to thoughtfully consider problems and issues, knowledge of logical and reasoning inquiry, and the ability to apply those methods" (p. 69). These broad definitions of critical thinking for the workplace and responsible citizenship will serve as an initial guide for selecting the skills and dispositions that will be assessed.

2. Higher Order Thinking Skills

Critical thinking skills are often referred to as "higher order cognitive skills" to differentiate them from simpler (i.e., lower order) thinking skills. Higher order skills are relatively complex, require judgment, analysis, and synthesis, and are not applied in a rote or mechanical manner. Higher order thinking is thinking that is reflective, sensitive to the context, and monitored. Computational arithmetic, for example, is not a higher order skill, even though it is an important skill, because it involves the rote application of well-learned rules with little concern for context or other variables that would affect the outcome. By contrast, deciding which of two information sources is more credible is a higher order cognitive skill because it is a judgment task in which the variables that affect credibility are multidimensional and change with the context.

Higher order cognitive skills are college-level critical thinking skills--the ones that are promoted in thoughtful college classrooms in which students grapple with complex issues and problems. They are skills that are (or should be) found in college-level texts, assignments, examinations, and some cocurricular activities. A national assessment of critical thinking skills should focus on such higher order cognitive skills.

3. Clear operational definitions

The critical thinking skills that are selected for assessment will go through several levels of review as the tests are designed, scoring criteria are developed, experts debate their meaning, and the general public considers the results of the assessment. It is essential that all skills have clear operational definitions that are commonly agreed upon and recognized by the developers and graders of the assessment instrument. Clear agreement will make scoring easier and less costly and will make communication about the process and the results more meaningful. The skills selected for assessment must also have "face validity," that is, they must look like reasonable and useful thinking skills to the general public who will be scrutinizing the assessment for its utility and cost-effectiveness.

Operational definitions are also required by the academic community whose job it is to enhance their development in college students. A critical reason for conducting a national assessment is that it will provide data-based recommendations to the decision makers in higher education. Faculty from every academic discipline should be able to identify those skills that they teach in their individual classrooms. A primary goal of the proposed assessment is the improvement of postsecondary teaching and learning; the skills need to be meaningful and useful to this important audience. If a national assessment of critical thinking is successful, then we should be able to draw diverse segments of the general public into the debate over what constitutes critical thinking and how we can promote its development at all levels of education. Thus, the skills that are selected and the language that we use in assessing them must be "sensible" to intelligent adults with no prior knowledge about this topic.

4. Cultural Fairness

The United States is one of the most culturally diverse countries in the world. All skills and all items that are designed to assess critical thinking skills must be scrutinized for possible bias. In order to provide a fair estimate of the ability of people from all cultures, the skills selected must be those that are used and needed in any cultural setting. For example, a frequently cited example of a question that is not culturally fair appears on the Weschler Intelligence Scale for Children (WISC). The question concerns what children should do if someone younger and smaller hits them. This question has been criticized by African-American psychologists who believe that "hitting back" (the wrong answer) is more acceptable in the African-American community than in other cultures because of the emphasis on caring for younger children. If the test developers had probed the thinking that went into the responses that children selected, they would have found this potential bias. The inclusion of questions like this one casts doubt on the validity of the WISC. This problem could have been avoided if greater concern for cultural fairness had been demonstrated at the time the test was written and at every stage of test development.

5. Multifaceted Nature of Thinking

The selection of critical thinking skills for the proposed assessment must also be guided by the belief that real-world thinking is a messy multifaceted enterprise. There is no single measure, no single question, no single skill that can capture the essence of how well an individual can think. A profile of skills across broadly defined categories will be a more accurate reflection of critical thinking than any single number summary.

A Taxonomy of Critical Thinking Skills

Using these criteria for the selection of critical thinking skills for a national assessment, a five-category critical thinking skills list is proposed. Skill level in each category can be assessed separately, in addition to an overall score. The categories and skills listed have face validity and, thus, can be easily communicated to the general public. Other taxonomies are possible and a legitimate case can be made for alternative groupings or adding other skills and deleting some of those that are listed. This is not meant to be the definitive list of critical thinking skills. Rather, it is proposed as a concrete starting place for the task of deciding what we want college graduates to know and be able to do so that they can compete and cooperate in the world's marketplace and function as effective citizens in a complex democratic community.

Here are the five category headings for organizing college-level critical thinking skills:

1. Verbal Reasoning Skills

The skills listed under this rubric include those that are needed to comprehend and defend against the persuasive techniques that are embedded in everyday language (also known as natural language). Thinking and language are closely tied constructs, and the skills listed in the section recognize the reciprocal relationship between language and thought, in which an individual's thoughts determine the language used to express them, and the language that is used shapes the thoughts.

2. Argument Analysis Skills

An argument is a set of statements with at least one conclusion and one reason that supports the conclusion. In real-life settings, arguments are complex, with reasons that run counter to the conclusion, stated and unstated assumptions, irrelevant information, and intermediate steps. Arguments are found in commercials, political speeches, textbooks, and anywhere else where reasons are presented in an attempt to get the reader or listener to believe that the conclusion is true. The skills of identifying conclusions, rating the quality of reasons, and determining the overall strength of an argument should be sharpened in college coursework.

3. Skills in Thinking as Hypothesis Testing

The rationale for this category is that much of our day-to-day thinking is like the scientific method of hypothesis testing. In many of our everyday interactions, people function like intuitive scientists in order to explain, predict, and control the events in their life. The skills used in thinking as hypothesis testing are the same ones that are used in scientific reasoning--the accumulation of observations, formulation of beliefs or hypotheses, and then using the information collected to decide if it confirms or disconfirms the hypotheses.

4. Using Likelihood and Uncertainty

Because very few events in life can be known with certainty, the correct use of probability and likelihood plays a critical role in almost every decision. Huff's (1954) tiny, popular book *How To Lie With Statistics* is still widely quoted because it explains how easy it is to mislead someone who does not understand basic concepts in probability. The critical thinking skills that are subsumed under this heading are an important dimension of a college-level critical thinking taxonomy.

5. Decision-Making and Problem-Solving Skills

In some sense, all of the critical thinking skills are used to make decisions and solve problems, but the ones that are included here involve the generation and selection of alternatives and judging among them. Many of these skills are especially useful in quantitative reasoning problems.

Taken together, these five categories (sometimes referred to as macro-abilities) define a workable model of critical thinking. They have the benefit of focusing on skills that are teachable and generalizable and would, therefore, help to bridge the gap between thinking skills that can be taught in college and those skills that are needed in the workplace. The critical thinking skills that are classified within these broad categories are presented in Table 1. An extended table that lists each skill with a description and an example of how it could be used is provided in Appendix A.

Table 1. College-Level Critical Thinking Skills* Categorization of Skills

1. Verbal Reasoning Skills
    a. recognizing and defending against the inappropriate use of emotional and misleading language (e.g., labeling, name calling, ambiguity, vagueness, hedging, euphemism, bureaucratese, and arguments by etymology [original word use]);
    b. detecting the misuse of definitions and reification;
    c. understanding the use of framing with leading questions, negation, and marked words to bias the reader;
    d. using analogies appropriately, which includes examining the nature of the similarity relationship and its connection to the conclusion;
    e. employing questioning and paraphrase as a skill for comprehension of text and oral language (i.e., recognizing main ideas);
    f. producing and using a graphic representation of information provided in prose form.

2. Argument Analysis Skills
    a. identifying premises (reasons), counterarguments, and conclusions;
    b. reasoning with "if, then" statements (which includes avoiding the fallacies of affirming the consequence and denying the antecedent);
    c. judging the credibility of an information source;
    d. judging the consistency, relevance to the conclusion, and adequacy in the way premises support a conclusion;
    e. understanding the differences among opinion, reasoned judgment, and fact;
    f. recognizing and avoiding common fallacies such as straw person, appeals to ignorance, slippery slope, false dichotomy, guilt by association, and arguments against the person.

3. Skills in Thinking as Hypothesis Testing
    a. recognizing the need for and using operational definitions;
    b. understanding the need to isolate and control variables in order to make strong causal claims;
    c. checking for adequate sample size and possible bias in sampling when a generalization is made;
    d. being able to describe the relationship between any two variables as positive, negative, or unrelated;
    e. understanding the limits of correlation reasoning;
    f. seeking converging evidence to increase confidence in a conclusion;
    g. considering the relative "badness" of different sorts of errors;

   h. solving problems with proportional and combinatorial (systematic combinations) reasoning;
   i. determining how self-fulfilling prophecies could be responsible for experimental results and everyday observations.

4. Using Likelihood and Uncertainty

   a. recognizing regression to the mean;
   b. understanding and avoiding conjunction errors;
   c. utilizing base rates to make predictions;
   d. understanding the limits of extrapolation;
   e. adjusting risk assessments to account for the cumulative nature of probabilistic events.

5. Decision-Making and Problem-Solving Skills

   a. listing alternatives and considering the pros and cons of each;
   b. restating the problem to consider different sorts of alternatives;
   c. recognizing the bias in hindsight analyses;
   d. seeking information to reduce uncertainty;
   e. recognizing decisions based on entrapment;
   f. producing graphs, diagrams, hierarchical trees, matrices, and models as solutions aids;
   g. understanding how world views can constrain the problem-solving process;
   h. using numerous strategies in solving problems including means-ends analysis, working backward, simplification, analogies, brainstorming, contradiction, and trial-and-error.

---

*Skills have been excerpted from Halpern (1989).

The Disposition To Think Critically

Critical thinking is more than the successful use of a particular skill in an appropriate context. It is also an attitude or disposition to recognize when a skill is needed and the willingness to apply it. There is an important distinction between what people can do and what they actually do in real-world contexts. This is called the competence-performance distinction. It is of no value to teach students the skills of critical thinking if they don't use them. A critical thinker will exhibit the following dispositions or attitudes:

1. Willingness to engage in and persist at a complex task

Critical thinking is hard work that requires diligent persistence, a trait that is central to academic success and most other success in life. In the jargon of cognitive psychology, tasks that require critical thinking have a high mental workload. A critical thinker must be willing to expend the mental energy that is required to begin and complete the task.

2. Willingness to plan

Psychologists call this the ability to inhibit action. A planful approach requires individuals to check their impulsivity and engage in intermediate tasks such as seeking additional information, organizing facts, and generating alternatives. The willingness to plan must become a habitual approach that is applied in many different contexts.

3. Flexibility or Open-mindedness

An attitude of flexibility is marked with the willingness to consider new options, try things a new way, reconsider old problems. It is the disposition to seek creative solutions. It is the antithesis of the rigidity and dogmatism that is characteristic of a "closed mind."

4. Willingness to Self-Correct

This refers to the willingness to learn from errors instead of becoming defensive about them. Individuals with the willingness to self-correct are able to utilize feedback and recognize the factors that led to the error. In order to improve the thinking process, ineffective strategies and automatic responses need to be recognized and abandoned.

5. Being mindful

Psychologists call this trait metacognition or metacognitive monitoring. It is the tendency to monitor one's comprehension and progress toward a goal. Critical thinkers develop the habit of self-conscious concern for and evaluation of the thinking process.

6. Consensus-seeking

Committee and group structures are most often the norm in the world of work. A critical thinker will need to be disposed to ways in which consensus can be achieved. Consensus-seekers will need high-level communication skills, but they will also need to find ways to compromise and to achieve agreement. Without this disposition, even the most brilliant thinkers will find that they cannot convert thoughts to actions.

Taken together, these six traits are characteristic of a critical thinking attitude or disposition. This concept is developed more fully in a tripartite theory of critical thinking dispositions, with dispositions categorized as inclinations, sensitivities, and abilities (Perkins, Jay & Tishman, in press).

Evidence That Higher Order Cognitive Skills Can Be Learned

Although there has been some debate about whether it is possible to produce long lasting enhancement in the ability to think effectively (Block 1985; Glaser 1984; Resnick 1983), there is a considerable body of evidence that critical thinking courses have positive effects that are transferable to a wide variety of situations. At least seven qualitatively different forms of outcome evaluations for thinking courses have been conducted, all of which are generally supportive of the idea that the ability to think critically can be improved as a result of thinking skills instruction. While none of these constitute the "perfect measure," taken together they provide a substantial body of support for the value of thinking skills instruction. (See Halpern, in press a and b, for an extensive review of these data.)

Data in support of the assertion that critical thinking is a learnable skill come from a variety of qualitatively different sources:

(1) A formal evaluation of a nationwide thinking skills program in Venezuela showed that students who had participated in classes designed for instruction in thinking skills showed greater gains in orally presented arguments and in answering open-ended essay questions than a comparable control group (Herrnstein, Nickerson, de Sanchez, & Swets 1986). This study is particularly notable because the oral arguments and writing samples were graded blind, that is, the graders did not know if the students they were assessing had received the thinking skills instruction or were in the control group. Results from

this project show that the targeted thinking skills were transferred and used appropriately with novel topics.

(2) Self-reports by college students show that the overwhelming majority believe that they have made substantial gains in their ability to think critically after completing a thinking skills course (e.g., Block 1985). Students rate themselves higher on numerous self-report scales including the willingness to suspend judgment, ability to evaluate conflicting claims, use of likelihood and uncertainty, and utilization of numerous problem-solving heuristics such as working backward from the goal, eliminating useless information, and evaluating the credibility of an information source.

(3) There is also research showing that college students who take a course in critical thinking show substantial gains in standardized tests of intelligence (Rubinstein 1980). Although there is ample reason to be skeptical about any claim to enhance intelligence that is attributable to a single course, this is an additional source of evidence that critical thinking courses can have positive effects.

(4) Other researchers have documented adult cognitive development that was enhanced with a college-level critical thinking course (Fox, Marsh, & Crandall 1983). Cognitive development was assessed with the Piagetian measures that have become standards for marking cognitive growth. This is an impressive improvement in light of the finding that only 25 percent to 50 percent of first-year college students possess the skills needed for abstract logical thought when this ability is assessed with Piagetian measures (Mc Kinnon & Renner 1971).

(5) A theoretically advanced means of assessing changes in thinking ability is to study the nature of the internal mental representation of a body of knowledge. Although the theoretical rationale for this technique is too technical for an extensive explanation in this context, the findings of these sorts of studies are easily comprehended. Schoenfeld and Hermann (1982), for example, found that when college students were taught general thinking skills, they demonstrated substantial improvement in problem solving relative to students who received more traditional instruction. In addition, these students organized their mental representation of the content matter in ways that were more similar to an expert's mental representation than those of a control group.

(6) Using a skills approach, Facione (1991) found that college students who received coursework in critical thinking scored significantly higher on a multiple-choice test of thinking skills than comparable students who had not taken such a course. The same finding was obtained with graduate-level college students by Lehman, Lempert, and Nisbett (1988) who developed their own multiple-choice test of thinking skills.

(7) Finally, Lehman and Nisbett (1990) examined the spontaneous transfer of thinking skills in an out-of-the-classroom, real-world environment. They phoned students at home several months after the completion of their coursework and posed questions under the

> guise of a household survey. Results were supportive of the idea that the students had learned and spontaneously used the thinking skills that had been taught in their college classes when the questions were asked in an ecologically valid setting (their own homes), with novel topics, several months after the semester had ended.

All of these diverse findings point to the same conclusion: college students can be taught to think more critically when they receive instruction that is specifically designed for this purpose. Ideally, different subsets of the skills of critical thinking would be reinforced and practiced in every college classroom, so that students learn to recognize and transfer skills across contexts. The assessment studies that used comparison groups, however, suggest that gains are most pronounced when instruction is specifically designed for the promotion of critical thinking. Critical thinking does not automatically result as a byproduct of standard instruction in a content area.

Section III. Resolving Critical Issues in Test Design and Administration

The need for providing information about the status of college-level critical thinking skills is relatively noncontroversial. There is currently little or no information to inform the decision makers (including students) in higher education, the workplace, the military, or in any other setting that is concerned with adult-level thinking skills. The controversies arise over questions of whether the information can be provided in a way that is meaningful, valid, useful, used, fair, and cost effective. A high quality assessment can provide many benefits to all of the other stakeholders in higher education and high level employment. But, if the assessment is not done well, the results will be costly and, possibly, harmful. It is clear that a national assessment of critical thinking skills must use valid indicators that are generalizable to all college graduates and must exercise caution to ensure that the data are not misused in ways that allow foolish and harmful comparisons. Public confidence and trust are the sine qua non for an undertaking of this significance.

Validity

There are many good texts on the psychometrics of test design and use. Interested readers are referred to the vast literature on this topic for definitions of terms and a review of basic concepts in testing such as reliability, effect size indicators, and measures of variability. For the purposes of this document, validity is the overarching psychometric concern. Discussion of other psychometric indices is not included. Readers who are unfamiliar with psychometric topics will want to consult some basic textbooks.

The focus of this assessment is college-level critical thinking skills. Based on the language in Goal 5.5, the college-level critical thinking skills that will be assessed are those cognitive skills that are needed for high-level employment and the thoughtful exercise of citizenship as well as the disposition to use these skills in appropriate contexts. A list of skills that meet these and additional criteria was presented in Table 1.

There are several ways of determining that a set of skills is a valid representation of the underlying construct (critical thinking) that is being measured. One of the most widely used means of construct validation is expert consensus. The skills presented in Table 1 are consistent with those in numerous sources that represent expert opinion, including leading textbooks that purport to enhance critical thinking, curricular objectives as stated in syllabi from a large sample of college-level critical thinking courses, Facione's (1991) report of the consensus obtained from recognized experts regarding critical thinking skills, documents written by the Critical Thinking Council (Paul & Nosich 1991), and scholarly work groups such as those sponsored by the American Psychological Association (APA) on undergraduate education and the American Association of College's (AAC) study-in-depth project. (The author was a participant in both of these work groups.) Based on the definitions and examples used in these various scholarly sources, the skills that are proposed have construct validity as determined by expert consensus.

As noted elsewhere, a national assessment of critical thinking also needs to have face validity; that is it has to "look like" critical thinking to the general public, in addition to having the more technical statistical properties that make it a valid measure. The skills listed in Table 1 have face validity in that they are categorized in a way that is similar to the language used in the Secretary's Commission on Achieving Necessary Skills (1992) and in Morse's paper on "Skills for Citizenship." Face validity is important because the proposed assessment should be a starting point for an inclusive national debate on critical thinking, that will be carried on in homes, schools, and industry settings. Ideally, newspaper accounts of the results will be followed by conversations about the nature of critical thinking that will help to establish criterion-referenced standards about the thinking skills that college graduates should know and be able to use.

Concurrent validity can be determined in the usual way by postulating and then verifying correlations with other instruments that would share some of the variance with a critical thinking assessment. For example, low positive correlations would be expected with the analytical reasoning portion of the Scholastic Aptitude Test, grades in critical thinking courses, the General Intellectual Skills Assessment in New Jersey, and selected portions of the National Adult Literacy Survey. These other measures, although not identical with the definition of critical thinking used in this paper, should have some overlap with the ability to think critically.

Data obtained from a national assessment of critical thinking will be valid only if the respondents take the test items seriously enough to demonstrate their capabilities. Motivating the test taker to perform at a high level is always a problem when an assessment is not directly tied to some outcome of consequence for the individual test taker. There are a variety of ways to make the results of the assessment of value to the individual. All of these methods involve "embedding" assessment into college practice. This means that assessment is not viewed as an "add-on" to the multitude of activities that are a legitimate part of college life. It must be routinized like other parts of the academic calendar (e.g., registration), used in ways that are meaningful to students, administrators, and faculty, and supported at all levels. A hard-learned lesson from educational assessment is the finding that assessment will not work if these criteria are not met.

First, provide feedback to the respondents that they can use in their career and course planning (for those enrolled in coursework). Make it clear that the assessment will yield valuable information that they can use to guide important life choices. If a repeated measures design is employed (as suggested below), information can be provided in ways that enable students to monitor their own cognitive and attitudinal gains (or losses) over time and relative to other normative samples.

Second, furnish individual information to the individuals' college and, if the person is employed at time of testing, to the employer. Such information will have to be provided in a way that is meaningful and fully acknowledges the limited role that such information should play in personnel and other decisions. It is important that the test takers know when they begin the assessment that their results will be forwarded to their college or employer.

A third approach is to use motivational techniques that have emerged as effective from psychological research. For example, charge a small fee for taking the test (people will take a task more seriously if they have paid to receive the results), explain that results will be used to compute group norms which will then be published, such as male and female norms, offer a reward (which would be underwritten by industry) for high performance such as a lottery ticket, free meals, or movies, or use game simulations to encourage respondents to demonstrate what they know. Motivational techniques are particularly important if the test format includes open-ended free-response questions in which the respondent has relatively few restraints and must compose a written or oral response. Questions of this type require more effort and, therefore, are more adversely affected by low motivation.

Unit of Analysis

A national assessment cannot succeed unless care is taken to ensure that the data are not misused. It would be foolish and dangerous to use the results to compare institutions that are dissimilar in fundamentally important ways such as the backgrounds of the students that they serve, their institutional missions, geography, funding, and faculty. Data need to be provided to the individual test takers to provide them with information and the incentive of personal accountability. Data should also be aggregated by type of institutions (2-year, 4-year, or doctoral/ research; private or public; and section of the country). Each institution can calculate or obtain a summary of the data obtained from its students or alumni, but such data about individual institutions should not be made available to other audiences. This practice would allow a college to judge how well it is performing relative to norms for similar institutions without allowing meaningless comparisons. I understand that this is a controversial proposal, but it is a workable compromise for resolving the uncomfortable tension between providing information about individual colleges that may promote foolish comparisons and providing information to individual colleges that can be used to gauge performance and guide reform.

Test and Question Format

A national assessment of critical thinking skills needs to balance several competing concerns. The ideal assessment would include several long-term observations, in real-life contexts, by multiple trained observers, with multiple-dependent measures. An assessment of this type would, of course, be prohibitively expensive and intrusive. A more realistic assessment that maintains some of these desirable properties is proposed. It includes a mix of questions, with a smaller number of costly, but valuable, open-ended questions and a larger number of questions that are amenable to machine grading such as multiple-choice and multiple-rating items. Secondary indicators such as portfolios, structured interviews, thought process protocols, and performance assessments are also suggested for a subset of the test takers. The secondary indicators offer a rich source of data that can serve as a useful guide to instruction and in the development of theories of adult cognition, but they are too costly and too unreliable to constitute a major portion of a large-scale assessment.

The Assessment

An ideal test will contain a mix of question types presented in a format that is easy to administer, store, and grade. A computer-assisted administration would allow for easy access in most colleges and workplaces, as long as security precautions are in place. A computer-administered assessment can also provide reaction time data, a major dependent measure in cognitive psychological studies. (Reaction times are sometimes used as measures of intelligence, and although it is not an appropriate use in this assessment, reaction times can help psychologists understand a great deal about the microcomponents of the underlying cognitive processes. They permit a much more fine-grained analysis of mental events than other commonly used dependent measures such as percentage of questions answered correctly. Reaction time data could be made available to cognitive psychologists for secondary analyses. They would not be provided to respondents or reported to the public.)

A good model question would have several sequential parts. The first part would require a short (e.g., two paragraph) constructed answer followed by more specific questions in a multiple-choice or multiple-rating format. The open-ended constructed portion of the question would allow graders to determine if the individual is predisposed to apply the appropriate critical-thinking skills. The multiple-choice or multiple-rating portion would show if the respondent is able to recognize the appropriate skill when it is presented in a list of alternatives. (Psychologists distinguish between recall and recognition paradigms. It is believed that they utilize different cognitive processes. Recall requires an extensive search of memory--it is tapped with the open-ended questions; recognition requires the verification of information that is presented to the individual--it is tapped with the multiple-choice or multiple-rating questions. Lower scores are expected in the constructed or free recall portion of the question than in the recognition portion.)

The two-part nature of the questions makes it possible to assess both the dispositional aspect of critical thinking and the ability to use a skill when a prompt is provided. The open-ended constructed response can be either written or oral. If both modes of responding are used (with oral presentations taped and written responses composed on the computer using a very basic word processor), a portion the communication goal of 5.5 can be assessed with the same instrument. (A written or oral response of this sort is one form of communication; it would not provide information about communication abilities in live, interactive settings which are arguably the most important arenas for high quality communication.) It is suggested that constructed responses be constrained in length to approximately two paragraphs so that scoring costs can be contained and good answers can be readily identified.

Although open-ended free response questions show how test takers approach a question and communicate knowledge, they can only be used to sample a few skills because they require a longer administration time than multiple-choice questions. They are also much more expensive to grade and lack the high level of reliability that can be obtained with carefully written multiple-choice questions. Multiple-choice responses can be machine graded at a relatively low cost. The multiple-choice portion of the assessment should also include multiple-ratings items in which there is more than one correct response or a judgment made about the quality of each alternative.

All questions need to be drawn from real-world contexts that are frequently encountered in the workplace and in the exercise of citizenship. This requirement will virtually ensure face validity and will be consistent with the "situated cognition" view point that is popular in the cognitive science literature (Glaser 1992; Rogoff & Lave 1984). Real-life thinking is done in a context, and a good test will provide a believable context for each question. Questions that are embedded in realistic contexts will provide a more valid index of abilities and improve motivation by engaging the interest of the test taker.

It is also suggested that two or three of the questions be presented in a video format because so much information is currently obtained with this medium. There are many video possibilities, for example, a hypothetical analysis of employment data presented in a news show context, a group of mid-level executives discussing ways to improve profits for a hypothetical company, or a physician explaining the risks and benefits associated with the use of a particular drug. Response format for the test taker can be the same as that used with written questions.

Several sample questions are presented in Appendix B. As seen in Appendix B, the nature of the proposed questions differ from those used on earlier assessment instruments in several ways. The questions should be relatively complex in that they contain several types of information. Like real-world thinking tasks, some of the information available may not be relevant, and part of the question involves deciding which information is important to the problem. What is important in an assessment of critical thinking skills is what test takers are required to do with the information. The tasks for the test taker should require thoughtful analysis and synthesis. Some suggestions for defining the task for the test taker are presented in Table 2.

Table 2. Examples of the task for test takers in an assessment of critical thinking skills.

After context-rich information is provided, the test taker is asked a question or told to perform a task. For example,

- Draw a diagram or other graphic display that organizes the information provided in the question.
- What additional information would you want before answering the question?
- Explain why you selected a particular multiple-choice alternative. Which alternative is second best? Why?
- State the problem.
- Which information is most important? Which information is least important? Why?
- Categorize the findings in a meaningful way.
- List two solutions for the problem.
- What is wrong with an assertion that was made in the question?
- Present two reasons that support the conclusion and two reasons that do not support the conclusion.
- Identify the type of persuasive technique that is used in the question. Is it valid or is it designed to mislead the reader? Explain your answer.
- What two actions would you take to improve the design of the study that was described?

Secondary Indicators

It is suggested that a national assessment of critical thinking be supplemented with a variety of secondary indicators. These indicators will add valuable information about the cognitive growth of adult students using ecologically valid indicators. The use of these indicators, in conjunction with the proposed assessment, can provide convergent validity for the assessment. If there is little agreement among these diverse methods of assessment, then results can highlight inconsistencies and suggest changes that can be made to improve the entire enterprise of determining how well college graduates can think.

Portfolios

Portfolios were originally developed as a means of assessing performance in the arts, but they can be used for many other purposes. A portfolio is a file of student-produced work that is periodically reviewed. There are numerous decisions about what to collect (e.g., students' best work, or typical work), when to add to the portfolio, and how much to include. Although reviewing a dated collection of student products can provide information about the student's progress, the problem still remains of how to assess the contents of the portfolio. How can the contents be summarized and communicated in a meaningful way?

Portfolios are a very costly technique with unknown reliability and validity, but they can suggest directions for improvement, and they can be a useful supplement in understanding the thinking and communication skills of students. Portfolios also present substantial storage and grading problems (especially if the products that are collected are not uniform). Portfolios can include video and audio tapes and other work samples in addition to written products. They have the advantage of face validity and could be used at selected colleges or work sites where there is faculty or employer interest in pursuing this type of project. It is suggested that several sites be selected to collect and examine portfolios, but that it be used as a supplement to a large-scale assessment.

Structured Interviews

Oral interviews also provide a rich source of information about critical thinking skills and, for this reason, are suggested as a secondary indicator. Like portfolios, they are very costly. Differences among the interviewers can also pose threats to their reliability and validity. The advantage of an interview is its close approximation to many real-world settings in which a test taker's response directs the nature of the next question or statement. Oral communication is an important component of an assessment of communication skills. For this reason, the structured interview would mesh well with the oral communication aspect of Goal 5.5. Costs can be reduced with group interviews, but groups add to the psychometric problems. On the other hand, group interviews can provide additional information that is important in a valid assessment because workplace thinking and speaking are often a group activity. It is suggested that selected sites be identified for the administration of a structured oral interview.

Thought-Process Protocols

In thought-process protocols, the respondents vocalize or write a description of their thought processes as they work through a problem or respond to a question. This technique can be very useful in examining the underlying thought process and in identifying gaps or peculiarities in the processing of information. It can provide useful data to faculty and researchers who are interested in education for thinking because it provides the best available "picture" of what the test taker is doing in the interval between the presentation of the information and the initiation of a response. A thought-process protocol can show if test takers have developed the metacognitive skill of monitoring their thinking process, a skill that is paramount in critical thinking but not assessed with the other measures. Thought-process protocols allow us to understand why a response was selected and why groups may differ in their question-to-question performance (e.g., gender or racial differences in responses to specific questions). It focuses more specifically on the thinking process than on the nature of the response.

Like the other supplemental techniques, the collection of thought-process protocols is costly and has limited generalizability. Nevertheless, it is useful for answering specific questions in education and in the formation of theories of adult cognitive development. They would provide valuable information about the thought process that precedes the response. Thought-process protocols are best used at selected sites where faculty or employers are knowledgeable about their use.

Performance Assessments

Performance assessments have an obvious advantage to pragmatists who argue that if you want to understand critical thinking, you must observe (or create) real-world settings where people are engaged in genuine tasks that require the use of critical thinking skills. It is a necessary environment for assessing the disposition to seek consensus that was described earlier. Despite the appeal of such an assessment, it too presents obstacles for generalizability, reliability, scoring, and the communication of results. Performance assessments are also an extremely costly technique. As a supplementary technique, they can provide information that cannot be obtained with the other techniques. Performance assessments can answer straightforward questions like "How do a group of college graduates solve a particular problem and how does their solution differ from a comparable group of adults who are not college graduates?" This is a technique that allows the investigator to examine individual or group problem solving in naturalistic settings and to devise tasks that would optimally require selected critical thinking skills. This would be a good supplemental technique for industry and business settings, although it can also be conducted in colleges.

Summary of Recommendations for Test and Question Format

A computer-assisted assessment with a mix of open-ended free response and multiple-choice questions is recommended. The questions will be relatively complex and will require critical thought in the production and selection of an acceptable answer. Ideally, some of the questions will be presented by video technology because this format is frequently used to communicate information in real-world settings. The open-ended free response questions will have several sequential parts--the first will be freely constructed so that test takers can communicate their knowledge and show if they are predisposed to use critical thinking skills. Additional parts will be machine-graded multiple-choice or multiple-response items in which test takers can show if they can recognize the best response when a cue is provided. A mix of written and oral responses to the open-ended free response portion will be recorded. Nontraditional supplemental indicators such as portfolios, structured interviews, thought process protocols, and performance assessments will be used on a limited basis to provide additional information.

Standards

> [S]tandards describe what optimal performance looks like and what is desirable for students to know (Office of Technology Assessment 1992, p.3).

There are two main types of standards--criterion-referenced and norm-referenced. Criterion-referenced standards are based on some expert's opinion about the level of achievement that should be expected from the test taker, in this case, a well-educated college graduate. Standards are often communicated with the designation of levels of proficiency. With criterion-referenced standards, it is theoretically possible for every test taker or for no test taker to achieve a given level of proficiency. By contrast, norm-referenced standards interpret scores as relative to some group norm. With norm-referenced standards, a score at or above a selected percentile (e.g., 75th percentile) might be labeled as "proficient." With this definition, scores are interpreted relative to other group members instead of a predetermined level of achievement.

Although criterion-referenced and norm-referenced standards may appear independent, in actuality they are frequently interdependent. It makes no sense to establish an arbitrary standard of proficiency before an assessment instrument is pilot tested. Normative data are needed to determine the level of achievement that is reasonably possible. It is suggested that proficiency standards be established after a pilot administration of the instrument so as to avoid meaningless conclusions such as every (or almost every) college graduate is a highly proficient critical thinker or none are.

It is also suggested that five separate scores be computed for every test taker because critical thinking is too multifaceted a concept to be captured in a single number. The five scores will correspond to the five categories of skills that were presented in Table 1 so that every report provides a critical thinking profile that can be used more effectively than a single number summary. The five categories suggested in Table 1 are verbal reasoning, argument analysis, hypothesis testing, using likelihood and uncertainty, and decision making and problem solving. The corresponding scores could be aggregated into a single score, if necessary. (I believe that the public will demand a single number.) Three levels of proficiency are suggested for each category and for overall assessment performance. The proposed proficiency levels are "below proficiency," "proficient," and "highly proficient." These terms are descriptive and meaningful to the general public.

There should be behaviorally anchored scale points for each level of proficiency for each category (e.g., to be proficient in argument analysis the test taker must be able to identify the conclusion and one reason in a two paragraph argument, judge the credibility of an information source in an appropriate manner, and recognize at least two common fallacies when they appear in everyday language). The exact points at which these levels are set need to be tied to realistic goals and at levels that are close enough to actual performance to bring about change. It is meaningless to establish numerical criteria without pilot testing an instrument, and I would resist setting the markings on a critical thinking yardstick in advance of any data collection. The points at which standards are set will be determined at a later stage in the development of the instrument.

Information about the disposition to use critical thinking skills will probably have to be provided on a separate scale. The details for scoring and reporting the dispositional data are not presented in this paper because they depend on too many variables that still have to be worked out, such as the number of open-ended free response questions, whether they are graded on a binary (used the skill spontaneously or didn't use the skill spontaneously) or multiple-point scale (e.g., used the skill well, average, poorly, or not at all), and whether it is possible to provide dispositional data for each of the five categories. It is virtually axiomatic that dispositions are more difficult to assess than skills. There have been several attempts to assess dispositions with self-report questionnaires, but the psychometric literature is replete with the problems of self reports, so I do not advocate their use. (Self-report data often bear little resemblance to reality.) The assessment of the disposition to use critical thinking skills is an important component of critical thinking, but a more detailed analysis of its use and means of reporting cannot be accomplished until a later stage of test development.

Communicating Results

The communication of test results needs as much planning and forethought as the other stages of test design. If the communication is done poorly, then all of the prior work is worthless. Individuals and other audiences need to be given information that they can understand and use. The data will not be used if results are cloaked in technological jargon that does not inform or if the process is perceived as invalid or "political" in a way that it is designed to provide benefits for some individuals or some colleges to the detriment of others.

The various audiences for the results have different needs, so no single report will work in all situations. Results need to be provided in a way that informs the individual test takers about how they are performing relative to others who are similar to them and relative to what is needed in the workplace and other contemporary settings. A five-part profile of results will provide information about the sorts of skills they still need to develop and those they are demonstrating at a high level. Institutions can also determine what aspects of their curriculum need to be strengthened and employers can know what sorts of skills they can expect from college graduates. (I believe that institutions will be more likely to thoughtfully consider the results and use the data to direct programmatic change when the data are not made public. Institutions whose students do not score well often exert considerable effort in discrediting the assessment and assigning blame when the results are available to the public.)

The issue of fairness--culture-fair, gender-fair, fairness to all socioeconomic groups--appears in several places in this paper. It is important enough to be mentioned again in this section on communicating results. It is impossible to anticipate all of the possible problems that can arise with respect to fairness. One way to minimize problems is to be certain to include, at every step of the assessment, a diverse group of reviewers and writers who are likely to be sensitive to the issue and to examine every item for possible bias. There are statistical techniques that should be used for identifying items that are being answered differently by individuals from different groups (e.g., gender differences, racial differences, geographical differences). The thought-process protocols can be useful in this regard for discovering why these differences are occurring.

Section IV. A Plan for Test Development

Although a carefully detailed plan for test development would go beyond the scope of this paper, a sketchy outline is proposed for those who have the responsibility for the next phase.

Staged Research and Development

The assessment instrument will need to proceed in a series of stages because development is a time-intensive, sequential process in which later decisions depend upon those that are made as the assessment program progresses.

A preliminary stage would involve writing and videotaping the assessment questions. The questions would be based on the thinking skills identified in Table 1 and would be pilot tested on small samples of high school seniors, college students, and individuals with and without college backgrounds in the workplace. Testing should take no more than 3 hours (including a 20-minute break) and must be completed at one sitting. Ideally, testing will be administered at a computer terminal and will require little or no assistance from a proctor. The video portion can also be controlled by the computer program along with instructions for responding. Graphs or other diagrams can be constructed with "mouse" devices that look and function like pens. Little or no experience with computers should be needed. (Research has shown that respondents prefer computer-assisted test administration to paper and pencil formats and that computer anxiety is not a significant variable in computer-assisted testing. Computer-adaptive testing, in which the computer program selects the next question for each test taker based on the answer provided to earlier questions, is a time- and cost-saving mode of testing, but it cannot be used in this context until we understand more about how individual's respond to specific test items.)

Test questions need to be distributed over the five categories that are proposed so that accurate information is obtained in each category.

Administration of Pilot Instrument

A 5-year test cycle is suggested for the administration of the pilot instrument. This cycle would allow for a longitudinal assessment so that changes over time can be tracked. If this plan is rejected because of its cost or lengthy developmental time span, the same sampling plan can used in a cross-sectional research design. I believe that there are many benefits to a repeated measures design (at least in part), and I hope that it will not be abandoned as too costly.

First Administration (Time 1)

It is proposed that the pilot be administered to a large representative sample of high school seniors in the second semester of their final year. Social security numbers will be used to identify respondents so that those who are retested at later administrations can be tracked. Realistically, testing in school settings is the cheapest and easiest way to obtain a large sample and should be used whenever feasible. The specific nature of the sample (e.g., sample size and selection of test takers) needs to be established by the testing professionals at a future time.

Second Administration (Time 2--2 years later)

Three groups of test takers will be sampled 2 years after the administration of the first test--(1) second semester sophomores in 2-year (community) colleges; (2) second semester sophomores in 4-year colleges; and (3) a sample of working and nonworking individuals who completed high school 2 years earlier but have not attended college. (Samples of unemployed can be obtained through employment offices, parenting associations, churches, and neighborhood groups.) Samples will be obtained from the military that fit into these groups. While an attempt will be made to sample many of the individuals who were in the first administration, new test takers will be added at this stage and many of the original test takers will be lost. (Repeated data analytic procedures will be used for those individuals who take the test more than once. Data obtained from test takers at a single administration will be analyzed with between-group data analytic procedures.)

Third Administration (Time 3--4 years after the first administration)

Three groups of individuals will be tested at this time--(1) students in the second half of their senior year of college (including students who transferred from 2-year colleges); (2) a comparable sample of same-aged working and unemployed adults who have fewer than one semester of college credits; and (3) a comparable sample of same-aged working and unemployed adults with 2 years of college credits completed, but less than 2 1/2 years of college credits (including those who graduated from 2-year colleges and those who dropped out of 4-year schools after obtaining 2 years of credit). Again, the military will also be utilized to provide appropriate samples.

Fourth Administration (Time 4--5 years after the first administration)

A sample of employed and unemployed adults with 4-year college degrees, 2 years of college credits completed, and less than one semester of college completed will be tested. This final testing of the pilot instrument will provide data regarding the employment status and critical thinking skills of adults 5 years after the completion of high school. It will allow us to determine the impact of college on critical thinking skills, employment status, and the relationship between level of employment soon after graduation and level of critical thinking skills.

The proposed experimental design will allow us to track changes in those individuals who were measured at more than one phase of the study and estimate the effect of repeated administration on test scores. It allows separate data analysis for 2-year and 4-year colleges and for students who transfer from 2-year colleges to 4-year colleges to complete their degree.

Simultaneous Administration of Skills Surveys

During the 5-year period in which the pilot instrument is being tested and refined, a large scale survey of students, employers, the military, and college faculty is proposed. The purpose of the survey is to identify those critical thinking skills that these constituencies believe are important to success in college, in the workplace, and in the home (including skills for effective citizenship). The list of skills that is presented in Table 1 will be presented to representatives of these constituencies so that each skill can be rated for its importance. Additional skills can also be added.

What We Can Learn From This Assessment

After 5 years, we will have a wealth of data about the levels of achievement that college graduates and adults with little or no college have achieved, in addition to data indicating which of the skills are believed to be most important for success in contemporary settings. By carefully specifying individual and institutional data elements, we can obtain preliminary answers to major questions in higher education, such as whether there is a relationship between the emphasis that a college places on teaching and the development of thinking skills in its students, the impact of the major on the development of critical thinking skills (e.g., Do science majors show a different pattern of skill proficiencies from other majors either before declaring their major or at graduation?), or whether sacrosanct variables like small class sizes and laboratory experiences have demonstrable effects on how well students think. By carefully collecting individual student data, statisticians could "hold constant" (via statistical procedures) potent variables such as family income level and level of achievement at the completion of high school. There are many potentially valuable relationships that could be examined empirically, some for the first time, with the proposed research design. A wealth of valuable information about teaching and learning could be provided to the higher education community.

The instrument will need considerable revision at end of the 5-year cycle with additional test administrations that extend beyond 1 year after graduation from college. International assessments of critical thinking could follow, if the preliminary results from the national sample show that such an assessment can provide valuable data. Although numerous benefits can accrue from the proposed assessment, the potentially most valuable outcome would be to prod the nation into a debate about critical thinking that would create a new emphasis on how well Americans think and on the effectiveness of college-level critical thinking instruction.

Section V. Caveats and Conclusions

I will leave cost estimates of a national assessment of critical thinking skills to professionals in the area of budgets. (I expect that no one will be able to provide reliable long-term cost estimates.) Just as I cannot estimate costs, I also cannot put a dollar amount on the potential gains that can be accrued from a national assessment of critical thinking skills. By planning a national assessment of college-level critical thinking skills, we are contemplating the grandest plan for understanding the thinking and learning skills of college-level adults ever undertaken. If the assessment is done well, it will provide an empirical basis for understanding what works in higher education.

There is a concluding step in this process that cannot be slighted. Thinking skills will improve only if the final loop of the assessment process is achieved--only if the data are used to improve teaching and learning. Followthrough will be needed at this point to ensure that the information provided is followed with action. There can be no skimping on the last phase, the one before the cycle begins again with curricular changes, test revision and administration, and so on. If there is not a commitment to the project from all participants to carry it through the action phase, there is little reason to begin the lengthy and costly process.

If the information is used to improve the thinking skills of college students, the potential gains are priceless. We have very good reasons to believe that a substantial portion of college students can improve their critical thinking skills when they receive instruction that is specifically designed for this purpose. As plans for a national assessment of college-level critical thinking skills move forward, faculty in every discipline should be encouraged to identify those critical thinking skills that they teach in their classes so that researchers can begin to identify where and how specific critical thinking skills are taught. For the first time, educational researchers will be able to link college-level instruction to outcomes. This would be a major step toward understanding how college students acquire and use information.

As America's best educated citizens become better thinkers, we should be able to ensure America's leading role in science and technology, boost the sagging economy, elect effective leaders who will protect democracy and freedom, and increase appreciation for literature and the arts. Today's college students will need the ability to think critically if we are to avoid nuclear war, conserve our natural resources, live peacefully with our neighbors on this ever shrinking planet, cure cancer, AIDS, and other dread diseases, find ways to feed the hungry, and eliminate a host of social problems ranging from child abuse to loneliness in old age.

Resistance and Realistic Expectations

It is hard to argue against the desirability of helping college students improve how they think, but every other aspect of the proposed national assessment can and will be debated. Virtually every plan for assessing educational outcomes is met with strong and vocal resistance. There are legitimate concerns that data will be misused, that the measurement of a construct as complex and multifaceted as critical thinking cannot be validly accomplished, that the cost of such an assessment will drain money from programs that provide the education that the assessment purports to measure, that critical thinking is too controversial and too value laden to be part of a national agenda, that this is the wrong time to begin a major undertaking, and more. It is true that all measurement is imperfect and cannot capture the essence of how thinking skills develop as a result of the college experience. But it is time to start an assessment anyway, because perfect measurement will never be possible. A response to all of the doubting Thomases was eloquently framed by Curry and Hager (1987, p. 57), "To assess outcomes, we must overcome enormous problems of procedure and analysis, but we cannot refuse to look at what the instruments enable us to see."

References

Alsalam, N., Ogle, L.T., Rogers, G.T., & Smith, T.M. (1992). *The Condition of Education 1992*. Washington, DC: U.S. Department of Education.

Bailey, T. (1991). Jobs of the future and the education they will require: Evidence from occupational forecasts. *Educational Researcher*, 20, 11-20.

Banta, T.W. (1991). *Toward a plan for using national assessment to ensure continuous improvement of higher education*. Washington, DC: National Center for Education Statistics.

Block, R.A. (1985). Education and thinking skills reconsidered. *American Psychologist*, 40, 574-575.

Brownstein, R. (1991, July 18). More conflicts likely in era of empty pockets. *Los Angeles Times*, p. A5.

Curry, W., & Hager, E. (1987). Assessing general education: Trenton State College. In D.F. Halpern (Ed.), *Student Outcomes Assessment: What Institutions Stand to Gain* (pp. 57-65). San Francisco: Jossey-Bass.

De Loughry, T.J. (1989, February 22). Report hits U.S. students' inability to grasp complex tasks. *The Chronicle of Higher Education*, 35, A2.

Facione, P. (1991, August). *Testing college-level critical thinking skills*. Paper presented at the 11th Annual International Conference on Critical Thinking and Educational Reform, Sonoma, CA.

Fox, L. S., Marsh, G., & Crandall, Jr., J. C. (1983, April 30). *The effect of college classroom experiences on formal operational thinking*. Paper presented at the 1983 Annual Convention of the Western Psychological Association, San Francisco, CA.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39, 93-104.

———. (1992). Expert knowledge and the processes of thinking. In D.F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 63-75). Hillsdale, NJ: Erlbaum.

Halpern, D.F. (1989). *Thought and Knowledge: An Introduction to Critical Thinking (2nd Edition)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. Publishers.

———. (in press, a). Assessing the effectiveness of critical thinking instruction. *Journal of General Education.*

———. (in press, b). Targeting outcomes: Covering your assessment needs and concerns. In T.V. McGovern (Ed.), *Handbook for Enhancing Undergraduate Education in Psychology.* Washington, DC: American Psychological Association.

Herrnstein, R.J., Nickerson, R.S., de Sanchez, M., & Swets, J.A. (1986). Teaching thinking skills. *American Psychologist,* 41, 1279-1289.

Huff, D. (1954). *How to lie with statistics.* New York: W.W. Norton, & Company.

Lapointe, A., Mead, N., & Phillips, G. (1989). *A world of differences.* Princeton, NJ: Educational Testing Service.

Lehman, D., Lempert, R.E., & Nisbett, R. (1988). The effects of graduate training on reasoning. *American Psychologist,* 43, 431-442.

Lehman, D., & Nisbett, R. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology,* 26, 952-960.

McKinnon, J.W., & Renner, J.W. (1971). Are colleges concerned with intellectual development? *American Journal of Psychology,* 39, 1047-1052.

Morse, S.W. (1989). *Renewing civic capacity: Preparing college students for service and citizenship.* 1989 ASHE-ERIC Higher Education Reports.

———. (1994). *Skills for citizenship.* Washington, DC: National Center for Education Statistics.

National Center for Excellence in Critical Thinking Instruction. (1991). Defining critical thinking. (Draft report). Santa Rosa, CA: Sonoma State University.

National Education Goals Panel. (1991). *The National Education Goals Report.* Washington, DC: U.S. Department of Education.

Nummedal, S. G. (1991). Designing a process to assess higher order thinking skills in college graduates: Issues of concern. Washington, DC: National Center for Education Statistics.

Office of Technology Assessment. (1992). *Testing in American Schools: Asking the Right Questions.* Washington, DC: U.S. Government Printing Office.

Paul, R.W., & Nosich, G.M. (1991). *A proposal for the National Assessment of Higher Order Thinking at the Community College, College, and University Levels.* Washington, DC: National Center for Education Statistics.

Perkins, D.N., Jay, E., & Tishman, S. (in press). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly.*

Resnick, L. B. (1983). Mathematics and science learning: A new conception. *Science,* 220, 477-478.

Rogoff, B., & Lave, J. (Eds.). (1984). *Everyday cognition: Its development in social context.* Cambridge: Harvard University Press.

Romer, R. (1991). Foreward. In National Education Goals Panel. *The National Education Goals Report.* (p.iii). Washington, DC: U.S. Department of Education.

Rubinstein, M. F. (1980). A decade of experience in teaching an interdisciplinary problem-solving course. In D. J. Tuma, & F. Reif (Eds.), *Problem solving and education: Issues in teaching and research.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc. Publishers.

Schoenfeld, A.H., & Hermann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 8, 484-494.

The Secretary's Commission on Achieving Necessary Skills. (April, 1992). *Learning a Living: A Blueprint for High Performance.* Washington, DC: U.S. Department of Labor.

U.S. Bureau of the Census (1991). *Statistical Abstract of the United States: 1991* (111th Ed.). Washington, DC: U.S. Department of Commerce.

Walberg, H.J. (1989). Science, mathematics, and national welfare: Retrospective and prospective achievements. In A.C. Purves (Ed.), *International comparisons and educational reform* (pp. 121-138). Alexandria, VA: Association for Supervision and Curricular Development.

Appendix A
Descriptions and Examples From the College-Level Thinking Skills Taxonomy

1. Verbal Reasoning Skills

category description: The skills listed under this rubric include those that are needed to comprehend and defend against the persuasive techniques that are embedded in everyday language (also known as natural language).

Skill	Description	Examples of Use
a. recognizing and defending against the inappropriate use of emotional and misleading language (e.g., labeling, name calling, ambiguity, vagueness, hedging, euphemism, bureaucratese, and arguments by etymology)	This is an assortment of common misleading verbal techniques based on language usage in which a bias for or against a position is created with the connotative meaning of the words used to describe and define the concepts.	Critical thinker (CT) recognizes the use of biased language in numerous contents such as the following examples: 1. accidental killing of US troops referred to as "friendly fire"; 2. use of labels such as "pro-choice" and "pro-life" to create favorable impressions; 3. report of research that "suggests" a finding instead of stating the results; 4. use of the term "disinformation" instead of "lies"; 5. calling an opponent a "pinko"; 6. obtaining agreement by stating "Every good American will agree that . . ."; 7. arguing that homosexuality must be sick because the word "gay" originally meant lewd and lascivious.
b. detecting misuse of definitions and reification	Both of these techniques are attempts to persuade--the first by assigning arbitrary definitions and the second by treating abstract concepts as though they had an objective reality.	In a discussion of whether alcoholism is a disease, CT knows that the answer depends on the way the term "disease" is defined. A hypothetical construct is defined and then treated as though it were objectively real, such as the use of terms like "self-defeating personality" to attribute blame to women who are battered.
c. understanding the use of framing with leading questions, negation, and marked words to bias the reader	A framed question, negation, or marked term creates an expectation for the type of response that is expected.	CT recognizes the bias in questions like, "Which of the presidential candidates is worse?" (implication that both are bad) and "Don't you agree that the company plan is sure to work?" (bias toward a positive response).
d. using analogies appropriately, which includes examining the nature of the similarity relationship and its connection to the conclusion	Analogies are effective persuasive techniques--good analogies are based on underlying relationships which are validly transferrable between the analogy and base domain.	In response to a suggestion that welfare recipients serve on welfare boards, a board member replied that this was like suggesting that the mentally retarded serve on the boards for their mental institutions. CT asks if the analogy between the mentally retarded (who cannot act intelligently) and welfare recipients is valid.
e. employing questioning and paraphrase as a skill for the comprehension of text and oral language	The ability to state a main idea and identify supplementary ideas is essential for comprehension.	An extended presentation is made in which the speaker includes a main idea, supplemental information, and irrelevant information. CT can summarize meaningfully what has been said.
f. producing and using a graphic representation of information provided in prose form	Converting information from verbal to graphic formats is one measure of comprehension that can be useful in applying the information provided.	Information about a spatial topic such as the plan for a new community is presented in prose. CT can convert the information to a spatial representation. Alternatively, CT can chart relationship information when it is presented in prose (e.g., family trees, hierarchies).

BEST COPY AVAILABLE

2. Argument Analysis Skills

category description: Argument analysis skills are those skills that are needed to judge how well evidence supports a conclusion. They involve considering counterevidence, stated and unstated assumptions, and the overall strength of the argument.

Skill	**Description**	**Examples of Use**
a. identifying premises (reasons), counterarguments and conclusions	An argument is an attempt to persuade the reader or listener with at least one reason and one conclusion.	An argument regarding the decision to increase social security will contain a conclusion, two reasons for the action, one reason against the action, and an irrelevant statement. CT will identify the parts of the arguments (but may not know the appropriate labels for these components).
b. reasoning with "if, then" statements (which includes avoiding the fallacies of affirming the consequence and denying the antecedent)	Conditional reasoning involves a contingency relationship such that if the antecedent is true, the consequence must be true. "If, then" statements are frequently misused in an attempt to mislead.	Speaker: "If the administration is doing a good job, then economic indicators will show improvement. Economic indicators are showing improvement, therefore, the administration is doing a good job." CT recognizes the fallacy in this type of statement.
c. judging the credibility of an information source	Judgments of credibility and bias are central to determining the quality of information. Credible sources have no basis for gain, appropriate expertise, and specific first-hand knowledge of the phenomenon.	The executives of a car company tout the safety features of their new model car. CT acknowledges the possibility of bias and seeks other sources of safety data.
d. judging the consistency, relevance to the conclusion, and adequacy in the way premises support a conclusion	The ability to analyze extended arguments that differ along several dimensions, including the overall strength of the conclusion, is essential in the workplace and for effective citizenship.	Information regarding a complex issue is provided, such as the decision to offer more choice in the selection of schools. Answers will be based on the information that is provided.
e. understanding the differences among opinion, reasoned judgment, and fact	Opinion is an unsupported preference, reasoned judgment is a conclusion with supporting reasons for believing that it is true, and facts have verifiable truth values.	CT recognizes that statements like "Unions are needed for the protection of workers, and that's a fact" are opinions. CT understands the difference between claims like "Sugar-O's are a good cereal because they have the RDA for fiber" and "I love the way Sugar-O's taste."
f. recognizing and avoiding common fallacies such as straw person, appeals to ignorance, slippery slope, false dichotomy, guilt by association, and arguments against the person	There are many common fallacies that need to be recognized--deliberately weak arguments, claims that something must be true because there are no disconfirming data, if X is allowed to happen, it will be followed by a similar but more drastic action, and Y cannot be true because of the person who supports it.	CT recognizes and avoids fallacies such as these: 1. We cannot change the immigration quota for one country because it will then be changed for all other countries. 2. The plan for the economy is bad because the third party candidate is advocating it. 3. I know that ghosts exist because no one has proven that they don't. 4. The new employee cannot be trusted because he dresses like a hippy.

3. Skills in Thinking as Hypothesis Testing

category description: The skills used in thinking as hypothesis testing are the same ones that are used in scientific reasoning--the accumulation of observations, formulation of beliefs or hypotheses, and then using the information collected to decide if the information collected confirms or disconfirms the hypotheses.

Skill	Description	Examples of Use
a. recognizing the need for and using operational definitions	An operational definition is an explicit set of procedures that specify how to recognize and measure a construct.	An advocate for a group claims that child abuse is increasing at an alarming rate in the U.S. CT will ask how child abuse has been defined and how it was measured.
b. understanding the need to isolate and control variables in order to make strong causal claims	In determining cause, a single variable is manipulated and the results attributed to that variable are compared to a comparable control group in which the variable was not manipulated.	A commercial states that cholesterol levels were reduced when a group began exercising and using margarine. It concludes that margarine use reduces cholesterol. CT notes the confounding of exercise and margarine use and does not attribute the drop in cholesterol to margarine.
c. checking for adequate sample size and possible bias in sampling when a generalization is made	Valid generalizations from samples can be made only when the sample size is adequately large (relative to the variability) and the sample is representative of the population.	As part of a conversation, a young adult states that she knows that old people like to watch Lawrence Welk because her grandmother did. CT recognizes that this sample is too small and biased for generalizations about all old people.
d.being able to describe the relationship between any two variables as positive, negative, or unrelated	Two variables are positively related when increases in one occur concomitantly with increases in the other, negatively when increases in one occur concomitantly with decreases in the other, and are unrelated when changes in one variable are independent of changes in the other.	A newspaper article states that over the last 10 years marijuana use has steadily increased and Scholastic Aptitude Test scores have steadily decreased. CT can describe this as a negative relationship.
e. understanding the limits of correlational reasoning	Although a significant correlation between two variables can suggest that changes in one variable cause change in the other variable, this is weak evidence for determining cause.	A social scientist shows that there has been steady increase in the number of single-parent families and in the number of crimes committed by juveniles over the last 15 years. She concludes that single-parent families are responsible for the increase in juvenile crime. CT notes that these data are correlational and cannot be used to determine cause.
f. seeking converging evidence to increase confidence in a conclusion	Multiple independent sources of evidence provide stronger support for a conclusion than single-source evidence.	The American Cancer Society advocates a low-fat diet to reduce the risk of cancer. This conclusion is based on evidence from other countries in which low-fat diets are associated with low cancer rates and high fat diets are associated with high cancer rates. In addition, experimental studies in the US show that cancer is reduced when dietary fat is reduced. CT recognizes that while each source of evidence provides weak support for the conclusion, taken together they strengthen the causal link between dietary fat and cancer.
g. considering the relative "badness" of different sorts of errors	Judgments about the consequences of different types of errors are important in determining how to act.	Mixed testimony is given in a trial. CT understands that it is worse to wrongly convict an innocent individual than to set a guilty individual free. (This is an essential component in our system of justice.)

BEST COPY AVAILABLE

h. solving problems with proportional and combinatorial reasoning	The ability to set up ratios that correctly express proportional relationships and to use a systematic approach to combinations is especially important in scientific and quantitative reasoning.	1. A forecaster provides information showing that when the government provided $2 million to colleges, the number of new jobs rose by 100,000. The government is now spending $3.5 million on colleges. If nothing else changed, how many new jobs would be expected? 2. Three candidates are expected to participate in a debate. The order in which they will be introduced has not been determined. How many different orders are possible?
i. determining how self-fulfilling prophecies could be responsible for experimental results and everyday problem solving	The term "self-fulfilling prophecies" refers to the conscious or unconscious tendency to act in ways that influence experimental procedures so that we obtain results that are consistent with our expectations.	A poll conducted by a conservative candidate shows that most people do not want to increase funding for the homeless. CT realizes that subtle effects in the way the questions are asked and nonverbal cues could be responsible for these results.

4. Using Likelihood and Uncertainty

category description: The correct use of objective and subjective estimates of probability is a critical thinking skill because virtually every life event and decision is probabilistic.

Skill	Description	Examples of Use
a. recognizing regression to the mean	An extreme score on some measure is most likely followed by a score that is closer to the mean.	It is a common phenomenon that a star "rookie" who excels in his or her first season performs closer to average in the second season. CT recognizes that this is an example of regression to the mean. (Some will recognize this as the "sophomore jinx.")
b. understanding and avoiding conjunction errors	The co-occurrence of two or more independent events is less likely than the occurrence of either one alone.	Physicians describe the typical heart attack victim as male and over 55. CT realizes that the typical heart attack victim is more likely to be either male or over 55 than to be both male and over 55.
c. utilizing base rates to make predictions	The initial or a priori proportion of some group in the population is a valid guide for predicting likelihoods.	A speaker describes an encounter at a food market with a "quiet man who is good with numbers." The speaker concludes that he is probably an accountant. CT knows that the number of accountants in the population is small relative to other occupations and prefers a more likely occupation such as teacher.
d. understanding the limits of extrapolation	Extrapolation is using trends in data to make estimations of future events, a process that is meaningful only if it is not extended too far in time and if other factors can be assumed to remain constant.	The population council concludes that, based on current birth rates, there will be no resources to feed the multitudes by the year 2050. CT knows that the extrapolation is based on the assumption that there will be no changes in contraception, fertility practices, or food resources.
e. adjusting risk assessments to account for the cumulative nature of probabilistic events	The probability of one or more unlikely events occurring increases with time and with the number of events.	A physician explains that there are 10 possible low probability side effects for a given drug. CT understands that although each side effect may be unlikely, given that there are 10 of them, the probability of any one of them occurring is additive.

5. Decision-Making and Problem-Solving Skills

category description: The skills used in decision making and problem solving are those involved in the generation and selection of alternatives and in judging among them. Many of these skills are especially useful in quantitative reasoning problems.

Skill	**Description**	**Examples of Use**
a. listing alternatives and considering the pros and cons of each	Every problem and decision involves selecting among alternatives. This is a systematic way to consider the advantages and disadvantages of various alternatives.	Several ways of increasing sales are described. CT considers the advantages and disadvantages of each and adds his own alternatives and combinations of alternatives.
b. restating the problem to consider different sorts of alternatives	Most real life problems are "fuzzy"; that is, there are many possible goals and ways to achieve them.	Several ways of increasing sales are described. CT redefines the problem as too little profit and considers other ways to increase profits.
c. recognizing the bias in hindsight analysis	Hindsight analysis is the re-evaluation of a decision after it has been made and its consequences known, with the belief that the consequences should have been known with greater certainty when the decision was made.	After a parolee goes on a killing spree, the townspeople want to fire the parole board. CT knows that the decision to release the parolee may have been reasonable given the information available at the time it was made.
d. seeking information to reduce uncertainty	Decisions based on more information are likely to be better than those made with greater uncertainty.	A company is deciding whether to increase its advertising budget. CT gathers relevant information about the effect of increased advertising before making the decision.
e. recognizing decisions based on entrapment	Entrapment is a situation in which much money, time, or effort have been invested and the decision to continue with a course of action is based on this investment.	The Pentagon argues that the government needs to continue spending money on a new weapon because it has already invested large sums of money on its development.
f. producing graphs, diagrams, hierarchical trees, matrices, and models as solution aids	Graphic representations of problems can be useful in solving them.	A problem is described verbally. The task for the CT is to depict the information in a graphic display in order to solve it.
g. understanding how world views can constrain the problem-solving process	There are limitations on the way individuals approach problems placed upon them by social class or other group membership.	A company president is confronted with a takeover. The possibility of cooperating with the competitor does not occur to the president because of the individual's world view.
h. using numerous strategies in solving problems,including means-ends analysis, working backwards, simplification, analogies, brain storming, contradiction, and trial-and-error	This is a collection of common strategies that every problem solver should know and use. They all require the planning and monitoring of a solution strategy.	Several stages of action need to be completed by a due date. CT works backward from the due date to decide how much time should be spent on each stage.

BEST COPY AVAILABLE

Appendix B

Sample Questions

A. A neighborhood watch program was initiated in the community of West Hills about 1 year ago. As part of that program, signs were posted to make visitors and residents aware of the program, and community education programs designed to help residents prevent crime were offered. The number of home burglaries has deceased substantially in this neighborhood over the last year. If you could ask two questions that would help you decide if this program reduced crime, what would they be?

1.

2.

Based on the information provided in this paragraph, what can you conclude about the effectiveness of this program in reducing crime? Please check one answer.

___ It is effective.
___ It is not effective.
___ There is not enough information provided to determine the effectiveness of this program.

[Answers would include the need for a comparison with a similar or nearby neighborhood. There are at least two other good possibilities: questions about whether the training was directed at reducing burglaries and the proportion of residents who participated in the training classes. There is not enough information provided to reach a conclusion.]

B. Last year, the Acme Widget Company had the worst business year in its 20-year history. Sales and profits were at their lowest levels ever for this company. In an effort to improve its financial status, management began a new "profit-improvement" program that included increasing the number of television advertisements for their widgets and laying off a large number of workers. Profits improved somewhat in the following year. Based on these facts, mark each of the following statements as either True (T), Probably True (PT), False (F), Probably False (PF), or Don't Know (DK).

1. ___ Increasing the number of television advertisements improved sales.

2. ___ Laying off a large number of workers improved profits.

3. ___ The combination of increasing television advertisements and layoff of a large number of workers improved profits.

4. ___ Profits would most likely have improved without any action because an extremely bad year is usually followed by a better year.

5. ___ In order to determine if increasing television advertisements improved profits, management would have to make a change in the number of advertisements while not changing any of the other variables.

[Answers would show an understanding of both confounded experimental designs and regression to the mean.]

C. A recent report showed that adolescents who smoke cigarettes also tend to get low grades in school. It found that as the number of cigarettes smoked each day increased, average grades in school were reduced. One suggestion that was made in this report is that we can improve school achievement by preventing adolescents from smoking. Based on this information, would you support this idea as a way of improving the school achievement of adolescents who smoke? Why or why not? Please list your reasons for your response.

Open-ended constructed response.

1. Based on the information provided, which is the best answer (check one).

___ We can conclude that one way to improve school achievement in adolescents who smoke is to prevent them from smoking.

___ We cannot conclude that one way to improve school achievement in adolescents who smoke is to prevent them from smoking.

___ School officials have no business getting involved in private matters such as whether adolescents smoke.

2. In no more than three sentences, explain your answer.

[Answer will show the distinction between correlation and cause. We cannot conclude that school improvement will result from preventing adolescents from smoking.]

ASSESSING THINKING:
A FRAMEWORK FOR MEASURING CRITICAL THINKING AND PROBLEM-SOLVING SKILLS AT THE COLLEGE LEVEL

by David Perkins, Eileen Jay, and Shari Tishman

Framing the Mission

How tall is a city? This is not an easy question to answer. Does one take as an index the tallest building or the average height of the buildings? Also, where does the city end? If we include the suburbs, the average height will be quite low. If we stick to the city center, the average height will be unfairly high. So where to draw the line?

Developing a framework for assessing people's critical thinking competencies poses similarly daunting puzzles. Good thinking is not one essence but a complex multifaceted enterprise—a structure of many buildings, so to speak. Some arguably are more central than others, but a one-dimensional model of thinking would lead to a single measure as misleading as a single measure of the height of a city. Worse, in the case of a city at least we know what different kinds of buildings there are. In the case of thinking, it is not so obvious what the components are—what kinds of thinking make up the whole.

In some sense, there can be no final answers to such issues. To make much progress at all, one must take into account the purpose of a model of thinking—what kind of work it is intended to do. In the current context, we are looking for a model that can best serve the needs of assessing college students' critical thinking and problem solving. What are these needs? Broadly, they are three in number and can be stated as principles.

Principle #1. An appropriate model of thinking on which to base a framework for assessment should be rich yet not unmanageably complex.

As suggested above, critical thinking is rich and multifaceted. For instance, solving a problem of formal logic is an example of critical thinking, but so is thinking broadly about the multiple points of view involved in solving a problem in the workplace. To be sure, both of these kinds of thinking processes can be called "critical," but they are really quite different. A model of thinking on which to base assessment must be rich enough to capture both, and other kinds of critical thinking as well.

Yet richness is a double-edged sword. Thinking is undeniably complex, but too complex a model will almost certainly create practical problems for an assessment initiative. For instance, it is simply unrealistic to expect that a critical thinking test that is practical in a college setting—one, say, that can be administered in 3 or 4 hours—can adequately measure large quantities of micro-skills across a sufficient number of dimensions and subject matters to be able to predict people's real-world critical thinking tendencies.

Principle #2. The model should identify thinking challenges and thinking behaviors that are relevant to real-world settings.

It is important to keep in mind *why* college students' critical thinking abilities are of current concern. As we enter the 21st century it is more urgent than ever to ensure that young people will be good critical thinkers and problem solvers in real-world settings—in the workplace, in the community, in their personal lives. All too often, assessments measure kinds of thinking that are relevant to a particular college course but are of little use beyond that setting. Yet, as their lives unfold, it is the settings outside of academia that will challenge young people's critical thinking, and in which they will enact their thinking tendencies.

Principle #3. A model of thinking on which to base a framework for assessment should indicate realistic norms for good thinking—norms that provide powerful yet achievable goals for instruction.

A model of thinking must do more than describe the nature or mechanism of thinking in general. It must provide a clear image of what *good* thinking is and how it happens, along with clear indicators for its measurement and improvement. In short, it must provide norms for good thinking. These norms should reflect high standards for thinking, of course, but they should also be realistically achievable in a college setting. That is, they should indicate standards that make sense in terms of instruction. Orienting instruction towards the standards embedded in the framework should improve students' critical thinking and their understanding of the subject matter.

The Plan of the Paper

In this paper, a model of thinking embodying the above principles is proposed. In the spirit of the NAEP initiative, this model can be used to assess how well college programs cultivate students' thinking. It can also be used to assess the critical thinking and problem solving of individual students. Here "critical thinking and problem solving" serves as a label for good critical and creative thinking in general, in keeping with the position of many participants in previous deliberations that there is not a hard line to be drawn between the two (Greenwood 1992). The paper presents an analysis of the kinds of thinking that might be measured and a sketch of an approach suitable for measuring those kinds of thinking. The model is supported by arguments for why we believe that this approach may be a useful one. We try to take into account many of the trenchant suggestions emerging from the previous conference and commissioned papers concerning the National Assessment of College Student Learning (Greenwood 1992).

A kind of fable introduces the approach proposed here, an idealized "fly-on-the-wall" story of assessing a person's thinking. This fable yields some key principles for assessment. Following a statement of these principles is the core section of the paper—a presentation of a model of thinking on which to base a framework for assessment. Rather than complicating matters by intertwining the presentation of the model with arguments for its defense, this section defers argumentation and describes the proposed model as simply as possible. The section that follows it provides a full technical rationale for each component of the model and addresses possible objections. Finally, the paper concludes with a section that briefly discusses some principles of assessment methodology and considers the practical implications of the proposed framework for college-based instruction. The appendix provides examples of sample assessments.

Fly-on-the-wall Assessment

Imagine a "fly on the wall" that follows a person about, observing what the person actually does. The fly reports back to an assessment team engaged in profiling people's thinking. Suppose one of the people is Susan, a student in her fourth year of studies in computer science, pondering a decision all students need to face around that time: What to do with the rest of her life, or at least, the first years of the rest of her life.

The fly hears Susan talking with her friend Carol about that very issue. Susan says, ''I have this offer from Exxon to work on a team that's centralizing all their data processing in one place, via optical cable and so on. It's pretty attractive, innovative work. But I don't know.''

"What's the problem?" Carol asks.

"Well, I made a big list," Susan says. "There are good points and bad points. Let me see how they seem to me now. Well, one good point that stands out is pay and another is that it's a pretty interesting project. But a bad point is I have to move to Texas and that's a long way from my family... And there are lots of other things."

"So it doesn't all balance out one way or another?"

"It balances out fairly positively. The trouble is, there are other things I could do, and they look good too."

"Like?"

"Well, like doing some graduate work. I haven't heard yet, but I might get into MIT for a master's program. And this guy I know is starting this company to design and build a cheap, very parallel supercomputer. That's risky but intriguing. And here's a crazy one I'm pushing around in my head. I've really tried to look around, you know, and find interesting directions. Well, there's some new work on biocomputing, ways to trick organic molecules into doing computation. I don't know much about it. But the future could really be there. I don't even know how to pursue it, but I'd like to find out if there's some program or company or something. That's a lot more exciting than the Exxon thing."

The Importance of Dispositions Alongside Abilities

The fly reports back to the assessment team, and they find reason to be pleased with Susan's performance. They note that in this real-life decision-making context Susan shows some important decision-making abilities. She can generate and think through a number of pros and cons about the course of action before her. She can conceive a number of courses of action worth considering, some of them quite different from others.

They note that also Susan shows some healthy thinking ''dispositions.'' That is, Susan is disposed to invest herself in important facets of the decision making process. She not only can but *does* think broadly about a wide range of options, list out pros and cons, ponder how they weigh against one another. Having the ability and strategies to do such things would help Susan little if she lacked the disposition to invest herself in what, after all, is a very fundamental life decision.

To honor this aspect of thinking, philosophers and psychologists have written about "thinking dispositions" in contrast with thinking abilities (e.g., Baron 1985; Dewey 1922; Ennis 1986; Passmore 1967; Paul 1990; Perkins 1992; and Perkins, Jay, & Tishman in press). A disposition is a behavioral tendency, for example the tendency to approach situations in an open-minded way or the tendency to be careful. Dispositions are crucial to thinking in real-life settings such as Susan's: A person may have the *ability* to think in an open-minded way, but not be disposed to do so. Simply having an ability does not guarantee that one will use it.

Not only is this common sense, but there is research evidence for it as well. For instance, our group conducted an investigation a number of years ago on people's tendency to look at both sides of a case. The study revealed that most people can readily generate arguments on the side of an issue opposite their own. They have the ability. But they tend not to explore the other side unless prompted. They lack the disposition (Perkins, Farady, & Bushey 1991). Baron (1991) and Kuhn (1991) report similar findings.

If thinking dispositions are so important, what exactly are they? The short answer is that they can be thought of as broad fundamental traits of mind, or broad behavioral tendencies in thinking. For example, the tendency to be open minded is a thinking disposition. So is the tendency to be intellectually careful, as is the tendency to be reflective or metacognitive. In a later section, we provide a list of seven key dispositions. For now, it is important to keep in mind that, like any broad behavioral trait, thinking dispositions are not monolithic in structure: several ingredients contribute to their expression in behavior. For example, the disposition to be broad and open-minded in one's thinking might depend on any number of the following factors:

- The habit of doing so
- The intellectual policy of doing so

- Intellectual standards related to the value of broad thinking
- Beliefs concerning the role of broad thinking in the pursuit of knowledge or options
- Confidence in one's ability to think broadly
- The commitment and courage to do so in the face of barriers

Clearly, dispositions as conceived above are crucial to good thinking. This suggests the following principle for assessment:

> *Principle #4:* Assessments designed to profile people's real-world thinking must identify both thinking abilities and thinking dispositions, and discriminate between them.

The Importance of Kinds of Thinking

Of course, dispositions and abilities don't tell the whole story of good thinking. Recall Susan's situation. The fly on the wall also noticed that she was engaged in a particular *kind* of thinking: decision making. This, too, seemed important. After all, there were other kinds of thinking in which Susan might engage: justification, explanation, problem solving, to name a few. Kinds of thinking are important to take into account when developing a profile of thinking because people's dispositions and abilities play out in different ways, depending on their understanding of the demands of the thinking situation they face. For instance, Susan might have the disposition and ability to think broadly in a decision-making situation because she sees it as an occasion to do so. But suppose she is faced with the task of justifying a belief or theory. Will she also think broadly when it comes to looking for evidence? Only the fly on the wall—or astute assessment techniques—will be able to tell us.

It might be observed that there are an infinite number of kinds of thinking occasions, with infinite variations in terms of setting, social dimension, and knowledge component. For instance, decision making might take place in the setting of the workplace or the family; it might involve collaborative or solo thinking; it might involve specialized knowledge or common sense. How can an assessment framework cope with such diversity? These factors--the setting of thinking, the social dimension, and the knowledge component--are certainly important, and a later section of this paper treats them in more detail. However, despite such differences, most thinking occasions fall into a quite limited number of basic types. What characterizes these basic types is their unique demand structures, and their ubiquity across domains and contexts. For instance, decision making occurs in civic life, in physics, in carpentry, in the workplace, at home. Despite these different settings, decision making in all of these areas share certain key demands. For example, it is almost always prudent in a decision-making situation to search for alternatives and reason thoroughly about options.

Another type of thinking that is basic is justification, because beliefs and theories need to be justified across all types of contexts in all types of settings, from specialized academic work, to personal relationships. And in most instances of justification, there are certain key demands--for instance the search for relevant evidence. Yet another basic type of thinking that has key demands across contexts is problem solving: Whether you are solving a chemistry problem or a problem of personal time management, you need to define the problem situation, identify your assumptions, and so on.

The fact that there are a few basic kinds of thinking, and that people's dispositions and abilities play out in terms of them, has important implications for the design of assessments that aim to profile real-world thinking. Stated as a principle:

> *Principle #5:* Assessments should profile how people's dispositions and abilities play out in terms of a limited number of basic kinds of thinking, such as decision making, problem solving, and justification, among others.

A Framework for Assessment

With all five principles in mind, we now turn to the presentation of a model of thinking on which to base a specific framework for what aspects of college students' thinking to assess (the same framework is equally relevant to lower levels of education, with appropriate adjustment of expectations). The model-based framework involves two levels of analysis of thinking: (1) general overarching dispositions and abilities, such as the disposition and the ability to think in a broad, adventurous way, or the disposition and the ability to think in a clear, careful way; and (2) the effective handling of important ''basic problem types'' such as decision making, explanation, and justification. The same assessments can be used to measure both levels.

We begin by introducing the approach and defining the two levels. Then some factors involving the roles of knowledge, collaboration, and setting in thinking are addressed. The presentation of the framework concludes with an outline of how assessments might be designed to measure both general dispositions and abilities and the handling of basic problem types at once. A technical justification for the details of the framework follows as a separate major section.

Defining an Approach

Any analysis of critical thinking and problem solving explicitly or implicitly reflects some model or conception of thinking—a vision of what good thinking is and how it happens. The same is true here. The framework we offer has at its heart a simple, two-part conception of how thinking happens. In a nutshell, this conception is as follows:

> *When people think, they (1) bring to bear general dispositions and abilities of thinking to (2) meet the demands of particular thinking situations (for example, making a decision, justifying a belief, or constructing an explanation).*

Consider Susan for example. Remember what the fly on the wall saw. Regarding general dispositions and abilities, Susan seemed to show some broad traits such as the disposition and ability to think imaginatively, and the disposition and ability to think strategically. Susan applied these traits to a particular kind of thinking: decision making about her career. She knew, for instance, that options should be sought with imagination. Of course, the assessors collating the data from the fly certainly would want more information, but the signs were promising.

In other words, we propose this two-part analysis of thinking because a close look at the way thinking works in the world suggests that both parts are real. On the one hand, people seem to have general dispositions and abilities. On the other hand, there are things to know about handling decision making specifically, and other kinds of thinking, that help in making the best use of these dispositions and abilities.

From the standpoint of assessment this conception of thinking generates two important questions for an assessment to probe:

1. What general dispositions and abilities does a person have available to help them in their thinking?
2. How well does a person bring those general dispositions and abilities to bear on particular types of problems, like making decisions, justifying beliefs, or constructing explanations?

Notice that these are different questions and both important. For instance, one person might be generally imaginative, but not realize that decision making calls for imagination in seeking options; the person might see decision making as a matter of choosing between the obvious options, rather than searching for less obvious alternatives. Another person might realize that decision making invites seeking new alternatives, but might lack skills of brainstorming or assumption/questioning that would help in this search. So to understand people's thinking, and to assess it, we need to pay heed both to general dispositions and abilities on the one hand and to how people apply them to particular types of problems on the other.

This does not mean that we need two different kinds of assessments, one for general dispositions and abilities and the other kind for different types of problems, like decision making. The same tasks can be used to answer both questions 1 and 2, as discussed later.

Seven Key Dispositions and Abilities

Upon hearing the fly's report, it seemed natural to conjecture that Susan's thinking would show broad traits such as an imaginative and strategic character. In general, critical thinking and problem solving are easily characterized as involving broad standards for certain kinds of intellectual behaviors that some people meet better than others. It is perfectly comprehensible, for instance, to speak of people as more or less planful, inquiring, intellectually careful, imaginative, and so on.

As emphasized earlier, we do not view such general traits as unitary and unanalyzable. Recall that a number of different factors appear to contribute to dispositions: habits, policies, explicitly held standards, confidence, courage, and so on. Moreover, certainly one can identify more specific dispositions and abilities of importance. The tendency to be intellectually careful, for example, may well include the more fine-grained ability to recognize logical fallacies, along with the disposition to use that ability to good effect. Nonetheless, it is important to recognize that fine-grained abilities and dispositions *when expressed in performance* seem to cluster into broader behavioral tendencies. That is, we tend to see micro-abilities such as fallacy recognition integrated into a behavioral tendency that also includes other related dispositions and abilities.

With this point in mind, we offer an analysis of critical thinking and problem solving as drawing on seven overarching "dispositions and abilities" (adapted from Perkins, Jay, & Tishman [in press]). These seven dispositions and abilities reflect a compilation and synthesis from a number of sources. They are listed and described in Figure 1. Ideally, good thinking includes all of these dispositions and abilities exhibited appropriately at different times depending on the thinking situation. While it could be argued that other dispositions and abilities that contribute to thinking exist, these seven have been selected to collectively represent the core of good thinking. The list of seven may not be final: additions, deletions, and regrouping are possible.

Our purpose in this section is to lay out a model of thinking as plainly as possible; this is not the place for a technical defense of it. In a later section we discuss the distinctive features of the total framework in detail, providing a rationale for each aspect and considering possible objections. For now, three key features of the dispositions and abilities model of thinking presented in figure 1 are worth noting:

1. There are a manageable number of categories. Most analyses of thinking include several dozen categories. We believe that a relatively simple scheme with less than 10 major categories is both psychologically and philosophically warranted and very important for the practicalities of assessment.

2. Dispositions and abilities are paired. Other models of thinking include both an abilities component and a dispositions component with no obvious relation between the categories. Again, this keeps the scheme simpler, a boon to assessment. We will argue later that it is psychologically and philosophically sound.

Seven Dispositions/Abilities for Good Thinking

1. *The disposition and ability to think in a broad and adventurous way.*
   The broad search for new ideas, imaginativeness and adventurousness in thinking, risk-taking, intellectual confidence and independence; the ability to generate diverse options and create new connections.

2. *The disposition and ability to sustain intellectual curiosity.*
   The tendency to wonder, probe, find problems, a zest for inquiry; an alertness for anomalies; the ability to observe closely and pose problems.

3. *The disposition and ability to think across multiple contexts and perspectives.*
   Flexibility; openness; the search for alternative frameworks, perspectives and views; the questioning of assumptions and transferring of concepts to new contexts.

4. *The disposition and ability to build knowledge and understandings.*
   The pursuit of connections, explanations, implications, consequences; an ability to build conceptualizations.

5. *The disposition/ability to be intellectually careful and clear.*
   Precision, thoroughness, alertness to possible error or inaccuracy; alertness to unclarity and need for focus; judiciousness, avoidance of hasty unwarranted conclusions.

6. *The disposition and ability to seek truth through reasons and evidence.*
   Questioning the given, demanding justification, seeking to disconfirm, fairness and impartial thinking, openness to other viewpoints.

7. *The disposition and ability to be metacognitive and strategic.*
   Self-monitoring and self-regulation of thinking; the application of standards to thinking; alertness to lack of direction, the use of strategies, the setting of goals, the making and execution of plans.

Figure 1

3. Dispositions are defined in behavioral terms, in terms of what people do (for instance, thinking broadly and adventurously). In many frameworks, dispositions are defined more as character traits (the person is creative, rather than thinks broadly and adventurously). Behavioral descriptions cleave closer to the needs of observation and assessment.

Earlier it was mentioned that the assessment framework proposed in this paper has two aspects. We have outlined the first aspect—a dispositions and abilities model of thinking-traits competencies. We now turn to a discussion of the second aspect, basic problem types.

Basic Problem Types

The fly on the wall reported on Susan's making of a decision. Naturally, Susan's decision-making situation has its own unique features. She was making a decision about which job

BEST COPY AVAILABLE

opportunity to pursue, and in doing so she draws on her specialized knowledge about the computer job market, and takes into consideration her own needs and interests. Nonetheless, some features of her situation are common to many decision-making situations. For example, she is faced with what seems like an either/or choice, and tries to think broadly about further options. Most decision-making situations are presented in an either/or fashion, and it is usually important to consider thinking more broadly. Also, Susan tries to identify pro and con reasons for a few promising options. Most decision points have pro and con reasons for various options, and decision-making is usually better when one vigorously probes for them.

Decision making can be called a ''basic problem type.'' In general, basic problem types are familiar kinds of thinking challenges that occur in many different settings across several domains and involve similar standards across those settings. Decision making is a basic problem type because people often have to think critically and carefully about decisions at work, at home, and in civic life, and usually do well to search broadly for options, carefully assess consequences, and so on. Likewise, justification is a basic problem type because people have to justify beliefs or theories in many different settings. And as with decision making there are general standards. For instance, it is almost always important to search evenhandedly for relevant evidence when trying to justify something, although what counts as good evidence may differ from context to context.

Of course, there are problem types within problem types. For instance, within the basic problem type of explanation there are styles of explanation such as causal explanation or functional explanation or explanation by constraint systems such as Newton's laws (cf. Collins and Ferguson, in press; Ohlsson, in press; Perkins 1992; Scheffler 1981). The purpose of the conception of basic problem types is not to identify specific instances of problem types or sub-problem types, but rather to draw attention to how specific instances are members of a larger problem-type species that has identifiable key features with generalizable standards of performance.

People's conceptions of basic problem types are important, because they affect which dispositions and abilities come into play. To appreciate the needs of decision making or justification, a person needs to recognize the central question of the problem type, the goal it addresses. Also, the person needs to honor the standards or criteria that govern good practice of the problem type.

With all this in mind, Figure 2 presents a list of three sample basic problem types, specifying their general objectives and some relevant standards, along with some further basic problem types mentioned in name alone. Unlike the list of seven dispositions and abilities proposed above, this list is not intended to be either comprehensive or exhaustive. There are many other problem types that could be added. The list is open ended and subject to revision. But a few centrally important problem types, such as those listed, would provide an adequate basis for assessment.

SAMPLE BASIC PROBLEM TYPES

Basic Problem Type: Decision Making

Central question: *What is the best decision, all things considered?*

Some generally applicable standards and the dispositions and abilities they most centrally call for:

- options, beyond the obvious (calls for broad adventurous thinking)
- evenhanded reasons, pro and con (calls for seeking reasons, thinking across multiple perspectives)
- weighing of all factors (calls for intellectually careful thinking)
- . . . others . . .

Basic Problem Type: Justification

Central question: *For what reasons is or should such and such be the case?*

Some generally applicable standards and the dispositions and abilities they most centrally call for:

- relevant and extensive evidence (calls for seeking reasons and evidence)
- appropriate argument structure (calls for intellectually careful thinking)
- multiple frames of reference and/or viewpoints (calls for open-minded thinking)
- . . . others . . .

Basic problem type: Explanation

Central question: *What explains this thing, event, situation, etc.?*

Some generally applicable standards and the dispositions and abilities they most centrally call for:

- relevant type of explanation — causal, principle-based, covering rule, etc. (calls for careful thinking and for building understanding)
- explanation implies or leads logically to thing explained (calls for careful and clear thinking)
- evidence for the correctness of the explanation (calls for seeking truth through reasons and evidence)
- . . . others . . .

Other basic problem types: Prediction, informal problem solving (everyday problems), formal problem solving (mathematically-based problems), planning, design (of artifacts and procedures), etc.

Figure 2

As with the dispositions and abilities, so with the basic problem types, it's worth highlighting a few helpful features of the framework that will be considered more carefully in the technical justification to follow.

1. The categories are relatively broad, keeping the scheme manageable. Thinking would be assessed category by category, with attention to the standards that go with each category. Certainly it is possible to produce a more detailed analysis of contributing microskills. However, as urged earlier, an efficient program of assessment cannot expect to address the multitude of microskills involved in a complex performance like thinking, any more than typical wide-scale assessments of writing do.

2. The basic problem types incorporate standards for various kinds of thinking. It is of course important that any analysis of thinking make a place for standards. We add that these standards should not be viewed as rigid; they require interpretation in context. For instance, although decision making generally calls for a creative search for options, there are situations in which the circumstances limit the options to two or three obvious ones.

3. The basic problem types capture *general* standards applicable across a variety of subject matters and settings. Of course, each of these subject matters and settings will exert its specific demands. The following section identifies some of these factors and addresses briefly how they might be taken into account.

Features of the Larger Setting

The model of thinking presented here has highlighted a conception of thinking emphasized at the outset: Thinking involves the application of general dispositions and abilities to particular thinking situations such as decision making. This conception motivates our two-level analysis. However, such a model cannot stand alone: It should be viewed against the backdrop of other features that make up the larger landscape of thinking. It's useful here to distinguish three such features:

Feature One: The setting of thinking. Thinking takes place in various settings, and different settings carry with them different standards, demands, and expectations. Important settings of thinking include the workplace, home and family life, civic life, and academic environments. While there is certainly overlap among them, each of these settings tends to have unique qualities that influence thinking and its evaluation. For example, problem solving in the workplace typically involves different expectations and demands than either problem solving in an academic setting or problem solving in a family context.

Feature Two: The social dimension of thinking. Thinking often occurs solo: People frequently (perhaps all too frequently) puzzle out their problems by themselves. But thinking also occurs in social contexts where people need to work together, honor and learn from one another's perspectives, and arrive at satisfactory mutual resolutions. Such situations invite dispositions and abilities specially adapted to a social context, and the assessment of thinking must take this into account.

Feature Three: The knowledge component. Many thinking situations depend on common or easily accessible knowledge. But sometimes thinking depends on highly specialized knowledge, as in physics, comparative literature, or chemistry. What kind and how much knowledge people have influences how and how well they think. This, too, must be considered when assessing thinking.

Clearly, it is impractical to assess college students' thinking across all combinations of the above features. Nonetheless, a program of assessment should address these features in some practical and appropriate manner. For example, suppose the aim is to assess students' dispositions and abilities concerning decision making in the workplace. In designing the assessment instrument and evaluating performance, one should keep in mind such matters as the demands and expectations of the setting, whether the decision making will occur solo or in collaboration with others, and whether it draws on specialized knowledge.

How Assessment Might Work

What makes this model appropriate as a framework for assessment? Recall the three principles identified in the Introduction.

1. A framework for assessment should capture the richness of good thinking yet avoid unmanageable complexity. The relatively small number of broad dispositions and abilities and basic problem types achieves this.

2. A framework for assessment should characterize real-life thinking behaviors and target real-life thinking challenges. Dispositions and abilities are expressive of common behavioral traits of mind; basic problem types are typical of the kinds of thinking challenges people frequently face in various dimensions of their lives.

3. A framework for assessment should indicate norms for thinking—yardsticks by which to measure and assess particular kinds of thinking. The intellectual standards embedded implicitly in the dispositions and abilities, and the standards explicit in the basic problem types, provide such norms.

What would one actually do to assess students, using this framework? For instance, would one have separate tests for each of the seven general dispositions, each of the seven general abilities, and all the basic problem types? No! This framework suggests a much more efficient approach.

Imagine Susan, our heroine of the fly-on-the-wall assessment, walking in to a more realistic assessment designed according to the framework described here. Susan might spend several hours working through several complex thinking tasks featuring a sample of basic problem types. For instance, the assessment might include two decision-making tasks, two justification tasks, and two explanation tasks, as well as further tasks addressing other basic problem types, with each task representing a somewhat different setting. Every task would require rich written responses, or might even involve interview techniques, depending on the data-gathering method employed.

How do we develop a profile of Susan's thinking from her responses?

- Scoring Susan's performances according to standards for decision making, justification, and explanation gives information on how well Susan handled those problem types. For instance, if Susan did particularly well according to the standards for decision making on both decision-making tasks and not so well on both explanation tasks, it's natural to conclude that she understands the demands of decision making better than those of explanation.

- Generalizing across Susan's performance on all three tasks tells us about her general dispositions and abilities. For instance, both the decision-making task and the justification task might call for broad adventurous thinking. If Susan showed such thinking on both tasks, this would be evidence that she has the general disposition and ability.

Methodologists will recognize that this sketch leaves many problems unanswered. To be sure how well Susan handles a basic problem type, one would need not one or two but several tasks representing the type, at least until validation studies had been done. To be sure about Susan's dispositions and abilities, one would need to generalize over a number of tasks—again, at least until validation studies had been done. Meanwhile, the basic point here remains that the same kinds of tasks—activities built around basic problem types—can yield information both about a person's performance on the problem types and about a person's general dispositions and abilities.

A practical program of assessment would work with all seven dispositions and abilities and a small selection of basic problem types deemed representative, perhaps half a dozen. Such a program of assessment also needs to respect the different contexts of thinking identified earlier—settings such as the home, the workplace, and civic life; the social dimension, thinking with others as well as by oneself; and the knowledge component, thinking that depends on general knowledge versus thinking in specialized contexts for which students may be especially prepared by their majors.

How far to go in exploring the combinations is a policy decision that must be made by those designing the program of assessment in light of the program's goals and considerations of feasibility. Clearly it would be impossible to assess students' thinking in all specialized contexts of knowledge, or in all social and solo thinking situations. Nonetheless, it is important for a program of assessment to bear these features in mind: Identifying them serves to locate assessment tasks in the larger landscape of thinking, thus indicating factors to be considered in the design and assessment of such tasks.

Rationale for the Framework: Six Useful Features

While the previous section outlined the basic structure of the proposed framework, this section provides the rationale for the framework. It focuses on in-depth justification for specific distinctive features of the framework and attempts to address potential concerns. Those readers who are interested in the technical justification will want to read this section carefully. Others more interested in practical application and assessment may want to skim this section and move on to the next section on assessment.

The model of thinking proposed here is firmly based on prior work. At the same time, it introduces several somewhat innovative features, unfamiliar to previous accounts of critical thinking. These features are especially useful for the twin jobs of guiding assessment and guiding instruction. The features are as follows:

1. A more comprehensive, two-part framework of thinking: Dispositions and abilities and basic problem types.
2. A manageable number of broad, encompassing categories.
3. Categories that work across different contexts of knowledge.
4. Dispositions defined in behavioral terms for easy identification.
5. Dispositions and abilities paired for structural simplicity.
6. Greater integration of standards into dispositions, abilities, and basic problem types to reflect their central role.

To justify the taxonomy proposed here, we review the rationale for these features and consider and rebut some possible concerns.

Feature 1: A More Comprehensive, Two-part Framework of Thinking: Dispositions and Abilities and Basic Problem Types

The proposed framework examines thinking and its assessment using two levels of analysis: dispositions and abilities on the one hand, and basic problem types on the other. Most taxonomies are lists of abilities and, in some cases, dispositions that contribute to critical thinking. So, why do we introduce in this framework an additional entity called basic problem types?

There are three important reasons. First of all, basic problem types have high face validity in academic, workplace, and citizenship settings. That is, they match the kinds of thinking episodes that occur in real settings. For example, people in many settings engage in episodes of thinking clearly focussed on decision making, explanation, and other problem types. Second, each basic problem type involves a fundamentally different core purpose and structure. Thus, problem types fall into distinct, natural kinds whose core structures are generally agreed upon. For instance, commonly accepted core structures are acknowledged for the construction of arguments (Beardsley 1975; Ennis 1969; Toulmin 1958; Toulmin, Rieke, & Janik 1979); dialogical reasoning, that is, thinking across frames of reference (Paul 1986, 1990); and design tasks (Goel & Pirolli 1992; Perkins 1986; Simon 1981). Third, basic problem types have cultural validity across multiple real-world contexts. As urged by Mentkowski (1991) in connection with the Alverno College program and by Nummedal (1991), thinking should be practical, paying off in multiple ways and contexts of life. So, while explanation, for example, takes somewhat different forms in anthropology, mathematics, comparative literature, or the workplace, there are discernable families and patterns of explanation and some common features of the explanatory enterprise.

With the role of basic problem types justified, we now turn to an examination of how the two-part framework achieves comprehensiveness—that is, why a conception of dispositions and abilities *and* a conception of basic problem types are both needed to tell the story of good thinking. Again, there are three reasons. First, basic problem types provide a context for dispositions and abilities; they are the arenas in which dispositions and abilities play themselves out. A thinking performance therefore expresses both the ongoing characteristics of the thinker, and his or her approach to the thinking situation at hand. Second, the conception of basic problem types helps explain how dispositions and abilities are activated in a given thinking situation, and how they cluster. For example, a rich conception of the demands of a justification situation appropriately cues and calls into play a cluster of dispositions and abilities such as seeking truth and evidence and looking at multiple frames of reference. Third, a framework that can identify the roles of dispositions and abilities, as well as basic problem types, allows for greater discrimination of errors in performance. For instance, it can detect whether a poor justification performance is due to a failure to recognize the thinking situation as a problem of justification, or as a shortfall in ability or disposition.

The above rationales notwithstanding, some concerns about a two-part proposal might still be raised. For example, it might be contended that existing taxonomies already include kinds of thinking like explanation and justification, so there is no need to create a separate category like basic problem types. Yet the above arguments indicate that combining abilities and dispositions with basic problem types is a category mistake. Basic problem types are simply not the same kind of thing as dispositions and abilities: Basic problem types are concerned with features of the thinking situation in which the thinker is engaged, while dispositions and abilities describe abiding characteristics of the thinker.

Another possible concern about the two-part framework might be that it is too complex. It is indeed complex, but so is thinking. We would caution against a simplistic approach to assessment that is not able to capture the genuine richness of real-world thinking performances. The challenge is to keep the complexity of the assessment methodology to a minimum, without sacrificing the integrity of the two-part model. We refer readers to the appendix for examples of relatively simple assessment methodologies based on the proposed framework.

Feature 2: A Manageable Number of Broad Encompassing Categories

All of the categories proposed in this taxonomy are deliberately quite broad. While dispositional categories usually are rather broad, this is not so commonly the case with categories of ability. For example, the ability to think in a broad adventurous way encompasses a number of subskills like the ability to brainstorm or the ability to identify and challenge assumptions. The ability to seek truth and evidence encompasses a number of subskills like articulating claims, identifying needed information, assessing the weight of evidence, and so on. The basic problem types likewise have an holistic character. For example, decision making involves subskills of generating options, foreseeing and weighing consequences, and balancing out factors to reach a decision.

In contrast, some taxonomies of thinking involve extended lists of quite specific kinds of thinking, such as syllogistic inference, probabilistic inference, hypothetico-deductive reasoning, differentiating truth of reasons from validity of deduction, and so on.

We believe that broad and encompassing categories serve better the enterprises of assessment and teaching. One reason is that there are simply fewer categories to cope with, at least at the top level of organization. Consequently, for the design of assessment and the design of instruction, it is easier to ''see the forest'' rather than getting lost among the individual "trees".

Another reason is that such broad and encompassing categories are truer to the functional organization of thinking. People do not engage in isolated acts of generating options or assessing consequences. They engage in decision making as a coordinated enterprise within which they generate options or assess consequences. Indeed, it is singularly uninteresting whether a person generates options or assesses consequences well in isolation, unless the person recruits those abilities in the service of thinking enterprises like decision making. Likewise, people do not engage in isolated acts of brainstorming or searching for questionable assumptions. They do so in the context of creative search, broad adventurous thinking after options in decision making, explanations in a quest for understanding, evidence in a quest for truth, or whatever the target may be. Thus, we understand a person's thinking ability better by looking at it in terms of broad functionally organized chunks than in terms of atomistic subskills.

A natural concern is that important subskills will get lost amidst this emphasis on larger categories. There are four somewhat different answers to this concern. First of all, the assessment of thinking as it plays out within basic problem types involves the assessment of subskills in the context of the problem type. Second, even if some subskills get lost, it is simply not practical to assess vast numbers of subskills independently of one another. Third, regarding instruction, the approach to teaching a performance by breaking it up into a host of subskills has not met with great educational success: It tends to be tedious and students tend not to integrate the subskills into holistic performances. Fourth, selected subskills of special importance certainly could be assessed in a targeted way. In that case, their contribution to thinking will be clearer if they are positioned as subcategories within a system involving a few large categories.

Feature 3: Categories That Work Across Different Contexts of Knowledge

Thinking occurs across different contexts of knowledge. In some cases it draws on rather general "common knowledge," while in other cases, it calls upon more specific knowledge within a discipline. In principle, it cannot be taken for granted that a given problem type, for example, decision making, has the same properties or involves the same abilities and dispositions across different contexts of knowledge, such as different academic disciplines or various everyday contexts. The dilemma is: to what extent are the categories of dispositions, abilities, and basic problem types to be considered general elements of thinking and to what extent are they context-specific?

Some scholars have argued that at least in certain ways, patterns of thinking such as evidential reasoning differ fundamentally across disciplines (e.g., McPeck 1981; Toulmin 1958; Toulmin, Rieke, & Janik 1979). Moreover, a considerable body of evidence in areas such as mathematics, physics, and medicine shows that especially skillful thinking is context-bound, highly dependent on a rich knowledge base in the domain in question (e.g., Chi, Glaser, & Rees 1982; Ericsson & Smith 1991; Glaser 1984; Rabinowitz & Glaser 1985).

Should we then conclude that good thinking is a "local" phenomenon, specific to the knowledge context in question? Should we surmise that Susan (in the fly-on-the-wall example), who thinks well about her career decision, may very well not think well about planning a vacation or designing a microcircuit? Not at all. The context specificity of thinking is a contemporary debate, not a settled issue. A number of scholars have challenged the arguments concerning domain specificity on a number of grounds (see e.g., Ennis 1989; Perkins & Salomon 1987, 1989; Salomon & Perkins 1989; Sternberg 1985). Without recreating the debate here, it seems most plausible to say that some aspects of good thinking are more general and some less general, that good thinking has its overarching dispositions and abilities and its more context-specific dispositions and abilities.

Similarly, basic problem types work at general as well as more context-specific levels. For example, the basic problem type of explanation can be applied in many instances, from everyday contexts like explaining why John has reason to be angry with Peter to disciplinary contexts like explaining why an author adopts a certain style. In its general form, explanation carries a core structure that is common for all cases. Within the category, there are also more specialized, sub-problem types such as explanation in science which involve more stringent standards and demands for evidence than the general form. There are even more specialized styles of explanation such as explanation with constraint systems (such as Ohm's law) that pose yet more specific standards.

Despite the strong face validity of the generality of dispositions, abilities, and basic problem types across contexts of thinking, it would be imprudent of any program of assessment to take this generality for granted. The degree to which competencies play out in a general fashion is an empirical question. An advantage of the proposed framework is that it allows for this empirical work while being useful regardless of the results. For instance, justification is important in science, in history, and in literature; assessors will want to know how students perform on justification problems in these areas regardless of whether their performances correlate across contexts or not. Similarly, the disposition and ability to be metacognitive are important in mathematics, in anthropology, and in economics; assessors will want data concerning them in all of these areas, regardless of whether students' metacognitive performance in one area is predictive of their performance in others.

We think it highly likely that empirical work will indicate some generality of dispositions and abilities and basic problem types across contexts. Nonetheless, the categories in the framework work well across different contexts of knowledge: The utility of the framework itself does not depend on obtaining positive data regarding generality.

Feature 4: Dispositions Defined in Behavioral Terms for Easy Identification

The seven dispositions are defined in terms of tendencies to exhibit certain kinds of behavior, for instance the disposition to think in a broad, adventurous way, or the disposition to seek reasons and evidence. In contrast, it's also possible to define particular dispositions in more attitudinal terms: for instance, having the creative spirit or respecting truth and evidence.

We believe that for present purposes behavioral phrasings have certain advantages. First of all, in their root meaning dispositions are behavioral: philosophically, a disposition by definition is a tendency for something to behave in a certain way. Second, assessment is well served by the more operational behavioral phrasings, which clarify what behaviors an assessment should test for. For example, compare the attitudinal "the creative spirit" with the behavioral "disposition to think in a broad and adventurous way," and compare the attitudinal "respecting truth and evidence" with the behavioral "the disposition to seek truth and evidence." The behavioral phrasings point to observable features of behavior.

For a third advantage, behavioral characterizations guide educational practice better by articulating more concretely the dispositions teachers need to cultivate in their students. To return to the same examples, if a teacher sets out to cultivate the creative spirit, this sounds nice but exactly what does it entail? However, if a teacher sets out to cultivate broad and adventurous thinking, a number of approaches come to mind: setting up situations where students are encouraged to think broadly and adventurously and rewarded for doing so, modeling broad and adventurous thinking in the classroom, marking out broad and adventurous thinking as one of several standards for good thinking, and so on.

One concern about behavioral characterizations of dispositions might be that important attitudinal aspects of the disposition get lost. To this we have two answers. First of all, we suggest that the attitudinal color of dispositions still shines through the behavioral descriptions. Consider again "the disposition to think in a broad and adventurous way" or "the disposition to seek truth and evidence." These do not sound like cold-blooded mechanical enterprises. Second, dispositions are not just attitudinal. Recall the earlier point that the disposition to behave in a certain way emanates from a combination of habits, policies, standards, values, confidence, commitment, and courage. Neither habits nor policies are attitudes.

Another concern might be that certain dispositions proposed in other taxonomies have no place in the scheme: They are somehow excluded by the behavioral cast of the characterizations. Where, for instance, would we place a disposition like intellectual courage or confidence in reason? We argue that such dispositions do fit in the scheme, once one unpacks their meanings. One might translate intellectual courage as something like the courage to maintain intellectual standards in the face of contrary pressures. Now we can see that attention to intellectual standards is a part of disposition 7 (metacognition). As to courage, this is one of the several factors that can underlie any disposition (habits, policies, standards, values, confidence, commitment, and courage).

Concerning confidence in reason, this could be elaborated as perceiving the value of reasoning and having confidence in one's ability to reason. Notice that both the perceiving value and having confidence are among the factors that underlie dispositions (again, habits, policies, standards, values, confidence, commitment, and courage). As to reasoning, the seven dispositions and abilities collectively characterize good reasoning. So to speak of confidence in reason is simply to speak of a trait implicit in every one of the seven dispositions identified.

Feature 5: Dispositions and Abilities Paired for Structural Simplicity

Another quite unusual feature of the proposed taxonomy is its pairing of dispositions and abilities. For instance, the taxonomy refers to "the disposition and ability to build knowledge and understandings" and "the disposition and ability to be metacognitive and strategic." To the best of our knowledge, all prior taxonomies have involved separate lists of dispositions and abilities with no obvious relation between the categories of dispositions and the categories of abilities.

Several factors recommend the pairing of dispositions and abilities. One is pragmatic: The result is a simpler more parsimonious taxonomy. Instead of thinking of several dispositions here and several abilities there, we can think of seven aspects of thinking, each with a dispositional side and an ability side. Such a taxonomy is easier to understand and communicate. It provides a more compact framework around which to design assessments and instruction.

Another factor is philosophical: Dispositions and abilities pair naturally. Remember, dispositions are most fundamentally behavioral tendencies (depending on several causes such as habits, policies, standards, etc.). It is logical to couple the tendency to strive for X with the ability to carry off X. For every disposition, one can speak of the ability needed to express it successfully. For every ability, one can speak of the disposition needed to mobilize it. Thus, to get our shoes tied in the morning, we need both the disposition (largely a matter of habit) and the ability (a perceptual-motor skill) to tie them. Likewise, to think in a broad adventurous way, we need both the disposition and the ability to do so.

A third factor is educational: Pairing dispositions and abilities helps to secure for dispositions a significant place in instruction. Since instruction tends to be abilities oriented, dispositions with entirely different names from abilities are easily ignored. Indeed, they sound like an additional and not well-understood instructional agenda. However, when dispositions and abilities are paired, cultivating dispositions becomes a natural and immediate partner to cultivating abilities. Teachers can build dispositions along with abilities by providing enough practice with abilities to build good intellectual habits; by rewarding and otherwise honoring the use of the abilities in question; and so on.

With these reasons in favor of pairing dispositions and abilities, what concerns might there be? Perhaps the most obvious concern is that historically dispositions and abilities have not been paired. This is indeed a reason for caution. There is often wisdom in prior practice. However, it is not a reason to hold back if something seems on analysis to be a genuinely good idea, which we argue is the case.

One can understand why the advantages of pairing dispositions and abilities have not been better recognized previously. For one factor, noted earlier, often dispositions are phrased in attitudinal ways. This makes a pairing with abilities much less natural. However, behavioral descriptions of dispositions, which we argued for in Feature #3, create the opportunity for a parsimonious pairing with abilities. For another factor, many taxonomies of thinking are strongly oriented toward rather specific subskills. Since thinking dispositions usually are seen as very broad, they do not pair naturally with subskills. However, the larger more encompassing categories of ability, which we argue for in Feature #2, create the opportunity to pair the resultant broad categories of ability with broad categories of dispositions.

A more specific concern about pairing dispositions and abilities is that the pairing might suggest that dispositions and abilities go hand in hand, for instance that to have the disposition to think in a broad and adventurous way means the same thing as to have the ability to think in a broad adventurous way. In reply, the phrase "the disposition and the ability" in no way implies that having the disposition is the same as having the ability. Indeed, the "and" in the middle signifies that both are needed and either the one or the other might be absent.

We note that whether in psychological fact dispositions and abilities tend to go hand in hand is an empirical question. A further advantage of the pairing is that it permits asking that question in a systematic way. In the kinds of assessments we envision, one would measure, for instance, people's disposition and people's ability to think in a broad adventurous way. By looking at correlations between dispositional and ability scores, one can gauge whether the disposition and the ability vary independently of one another or together or somewhere between. With separate lists of dispositions and abilities, how the two relate becomes a much harder question to answer.

One last concern about pairing dispositions and abilities is that the pairing might constrain the list and squeeze out important dispositions on the one hand or important abilities on the other. A partial answer to this concern has already been given. We noted earlier that the present taxonomy makes room for dispositions described in attitudinal ways—once they are translated into more behavioral language. Moreover, we noted that many skills not named in the present taxonomy are simply subskills that fit within one or another of its categories.

A more general and philosophical answer returns to the point made earlier that a pairing of dispositions and abilities is inherent in the logic of the concepts. In particular, if a disposition is described in behavioral terms, of necessity one can then speak of the abilities needed to exhibit the behaviors in question. Likewise, given any ability, one can speak of the disposition to mobilize that ability. In other words, as a point of logic, pairing dispositions and abilities *cannot* exclude any candidate disposition on the one hand or ability on the other hand. To be sure, in crafting the proposed taxonomy of dispositions and abilities, we may have left out an important category, although we have tried not to. But, if we have, this is not because the pairing forced us to but because we simply missed another pair we should have included.

Feature 6: Greater Integration of Standards into Dispositions, Abilities, and Basic Problem Types to Reflect Their Central Role.

Intellectual standards are an important feature of thinking. Intellectual standards function as guideposts for the regulation of our thinking. In some taxonomies, intellectual standards appear as a separate list. However, in the taxonomy proposed here we have assimilated intellectual standards into the list of dispositions and abilities and have incorporated standards as a central feature of the structure of basic problem types.

In the case of the dispositions and abilities list, intellectual standards are assimilated into the list and serve as the underpinning for a number of dispositions and abilities. For example, the standards of clarity and precision are reflected in number five. The standard of evidence is reflected in number six.

We think that this move serves the assessment agenda well. The aim of assessment is not to measure standards per se, but whether people hold standards, not just in the sense of lip service but in the sense of actually behaving in accordance with such standards. In other words, one wants to measure people's behavioral tendencies—their dispositions. Of course, a disposition depends on more than standards; as underscored several times, it depends on habits, policies, standards, values, and so on. Nonetheless, standards are included.

A possible reservation is that standards are not sufficiently honored by bundling them into the list of dispositions and abilities. In reply, standards are honored in three separate ways in the proposed framework. First of all, specific terms like clarity, evidence, and so forth, relating to standards appear throughout the categories of dispositions and abilities. Second, the seventh disposition and ability makes specific reference to observing standards. Third, as just noted, the importance of standards is specifically acknowledged among the factors underlying dispositions (habits, policies, standards, etc.).

A related reservation might be that for teaching purposes it's useful to point specifically to a list of intellectual standards. We agree with this. We note that the proposed list of seven can be used as a list of standards, in addition to a list of dispositions and of abilities. That is, each item is "standard-like," it defines a kind of behavior to be sought after in the quest for good reasoning. One could even call the list a list of dispositions and of abilities and of standards, although this would be awkward.

In the case of basic problem types, standards play an important, explicit role because they are worked into the specified core structure for each problem type. Recall that the core features of a given problem type are represented in a central question that frames the goal of the problem type and in supporting factors that work toward achieving the goal. In our framework, the core features are viewed as standards or norms for performance within the problem type. We argue that the core structures for a given problem type set the standard or ideal behavioral elements necessary to qualify as good decision making, explanation, or whatever. Without established standards and core structures it is impossible to tell whether an individual is truly engaged in the essential elements of a particular problem type. For example, one would not be truly engaged in justification if one's thinking performance did not include appropriate argument structures or use of relevant evidence. We further argue that these core structures are normative because, while not always articulated, they are implicitly shared conceptions of what constitutes the structure of a particular problem type.

For the purposes of assessment, standards for problem types serve two important functions. They help to define what students' conception of a given problem type should be. And they also sets standards for working through the problem type effectively: Each of the supporting factors in the core structure is an index along which performance can be rated.

A possible concern might be that specifying the core structure for each problem type is too fixed and recipe-like. However, viewing these core structures as standards for performance means that the thinker is engaged in the structure sufficiently to satisfy the essence of the problem type. It does not mean that the individual must cover every subfactor in a rigid way. Indeed, the core structures are to be viewed truly as standards or norms rather than as absolute requirements.

Three Principles for an Assessment Methodology

The presentation of the proposed framework for assessing college students' critical thinking is now almost complete. In accordance with the two principles mentioned earlier, we have proposed and provided a rationale for a framework that can (1) identify both thinking abilities and thinking dispositions and discriminate between them; and (2) profile how people's dispositions play out in terms of basic problem types.

Yet a final task remains. We have proposed *what* should be assessed—thinking dispositions and abilities, and basic problem types. Yet embedded in this proposal are also implications for the "how" of assessment—for principles underlying *how* thinking should be assessed.

What follows, then, are three principles for a methodology of assessment. In an effort to provide a single list of principles for the entire proposed framework, these principles are numbered #6, #7, and #8 to indicate that they follow the principles discussed earlier. See Figure 3 for a full list of all principles.

Use Authentic Assessments

> *Principle #6.* Thinking dispositions and abilities and basic problem types should be assessed through ''authentic tasks,'' rich simulations of situations that demand thinking. Performance should be gauged in terms of criteria relevant to real-world performance.

Authentic assessment rightfully has become one of the rallying cries of innovation in testing (Gifford & O'Connor 1991; Perkins 1992; Perrone 1991; Schwartz & Viator 1990). The general idea behind authentic assessment is that the tasks posed should have authenticity as performances in the contexts in question. For example, if the aim is to assess people's decisionmaking, test-takers should encounter fairly realistic decision situations in academic, job-related, or personal contexts. The criteria applied should have transparent relevance to the demands of real-world performance. A good litmus of authenticity is the question: In preparing students for citizenship and the workplace, does it make sense to teach to this test? An affirmative answer signals that the test simulates full rich performances of the sort for which the instruction aims to—or should aim to—prepare students.

Other kinds of testing can of course yield valuable information about specific skills. Artfully designed multiple choice items can probe some aspects of good thinking (cf. Norris and Ennis 1989; Norris 1991). Recently, Scriven (1991b) has argued the virtues of multiple rating items, which involve not just selecting from response options but rating them in various ways. Paul and Nosich (1991) provide examples of multiple-choice and multiple-rating items as well as of more open-response items.

Nonetheless, both multiple-choice and multiple-rating items place the test-taker in a position of making discriminations about given stimuli, rather than constructing complex, composed responses, which is the inherent demand of most workplace, citizenship, and academic thinking situations. Such styles of testing may add significantly to the assessment of thinking, pinpointing nuances of skill hard to detect with authentic assessments. But they should not in our view be the mainstay.

The sample assessments in the appendix of this paper are examples of the kinds of authentic tasks enjoined here.

Probe Thinking Processes, Not Just Products

> *Principle #7.* Thinking should be assessed with techniques that probe the process of thinking, not just the final products of a course of thought.

Remember the fly-on-the-wall story? The fly catches Susan in the midst of a decision, not with one fully worked out. The assessors, hearing the fly's report, discover a window into Susan's thinking process. They hear about options under consideration, pros and cons weighed already and to be weighed again. Imagine how misleadingly stark and simple a picture would emerge if Susan simply were asked after she had made up her mind what her rationale was.

From the standpoint of assessing thinking, the final products of thinking are overpurified. Assuming a reasonable outcome (e.g., a good decision), the quality of thinking has at least as much to do with the process that went into it as it does the product. The quality of decision making, for example, has to do not just with the final decision taken but the number, variety, and promise of the options explored, the care with which consequences are explored and appraised, and so on.

Sensible as this is, what to do about it is not obvious. How can a thought process be tracked? How can one know what a person is thinking? Even the fly on the wall's report is partial, taking advantage of what Susan happens to mention to her friend.

The challenge is real but not insurmountable. At least three different techniques offer some entry into a person's thinking. Perhaps the best known is the collection of think-aloud protocols (e.g., Ericsson & Simon 1984; Newell & Simon 1972; Perkins 1981). A test-taker thinks aloud into a tape recorder while addressing a task. While this may be too unwieldy for large-scale testing, a more streamlined variant is thinking on paper. The test-taker is asked to write down his or her thinking along the way through the task. To be sure, what gets written down will only sample what was thought. Still, it adds to the information. Yet a third approach is stimulated recall. Here, immediately after completing a task, the test-taker answers a number of probing questions designed to elicit the pattern of thinking.

All of these have been used. All have their methods, hazards, and tradeoffs. Perhaps the greatest issue concerns whether data gathered in such ways distort the very process of thought they are intended to probe (Nisbett & Wilson 1977). In our view, the answers to this and other concerns are satisfactory enough to warrant the use of one or another technique (Ericsson & Simon 1984). The alternative--judging a person's thinking from the final outcome--is simply too precarious. Even if subject to a degree of error, methods that probe process would serve better.

Sample Different Contexts and Settings of Knowledge

> *Principle #8.* Assessments of thinking should sample dispositions and abilities and basic problem types in different contexts of knowledge, and should draw on various features of the larger setting of thinking.

The fly-on-the-wall story throws into relief another dilemma of assessment. Remember that in Susan's case, the fly saw only one situation of thinking, a personal career decision in computer science. In principle, it cannot be taken for granted that Susan will be good at decision making across different contexts of knowledge, and in settings with different features.

Therefore, to begin with, assessments should be designed to sample different contexts of knowledge, including relevant academic disciplines as well as "common knowledge" contexts. (See the first sample assessment in the appendix for an example of a common knowledge assessment task.) This applies to the assessment of basic problem types as well as the assessment of dispositions and abilities. For example, students might be asked to think through a problem of justification in science, a problem of justification in art history, and a problem of justification in everyday civic life. In all cases, certain competencies, such as the disposition and ability to seek truth and evidence, will be at a premium, and assessments will need to identify whether students transfer such competencies across contexts.

Further, assessment tasks should sample additional features of the larger setting of thinking. For example, real-world thinking often involves collaboration. Assessments therefore should occasionally require students to work in pairs or groups, and should be evaluated accordingly. Also, some assessment tasks should be designed to reflect the demands and expectations of environments other than academia, for instance the workplace, personal and family life, and civic life.

At the same time that one wants to sample various thinking situations, it is important to remember that it has been argued that it is reasonable to expect some generality in students' dispositions and abilities and in their approach to basic problem types. Only empirical work will show the extent to which students' competencies transfer across occasions of thinking.

Eight Principles for the Assessment of Thinking

1. **Use a rich model.** An appropriate model of thinking on which to base a framework for assessment should be rich yet not unmanageably complex.
2. **Make sure the model is relevant to real-world settings.** The model should identify thinking challenges and thinking behaviors that are relevant to real-world settings.
3. **Use a model that indicates appropriate norms.** A model of thinking on which to base a framework for assessment should indicate realistic norms for good thinking — norms that provide powerful yet achievable goals for instruction.
4. **Assess overarching dispositions and abilities.** Assessments designed to profile people's real-world thinking must identify both thinking abilities and thinking dispositions, and discriminate between them.
5. **Assess basic problem types.** Assessments should profile how people's dispositions play out in terms of a limited number of basic problem types, such as decision making and justification, among others.
6. **Use authentic assessments.** Thinking should be assessed through "authentic tasks," rich simulations of situations that demand thinking. Performance should be gauged in terms of criteria relevant to real-world performance.
7. **Probe thinking processes, not just products.** Thinking should be assessed with techniques that probe the process of thinking, not just the final products of a course of thought.
8. **Sample different contexts and settings of knowledge.** Assessments of thinking should sample dispositions and abilities and basic problem types in different contexts of knowledge, and should draw on various features of the larger setting of thinking.

Figure 3

But, as has already been shown in a previous section, it is reasonable to expect that a reliable profile of students' real-world thinking competencies can be drawn from a handful of carefully selected assessment tasks—tasks that judiciously sample from different contexts, settings and dimensions of thinking.

Summarized in Figure 3, the set of the three principles sketched above together with the five principles identified earlier suggest a comprehensive approach to the "what" and "how" of assessment. They identify what kinds of thinking should be assessed, and prescribe appropriate guidelines for how assessment should proceed.

Teaching to Tests Worth Teaching to: An Outlook for Education

"Education is test driven!" This is a commonly heard lament, and rightly so, when the tests that drive education typically measure the quantity of facts students possess rather than students' potential for real-world critical and creative thinking. However, test-driven education need not be a bad thing, so long as the assessment framework that drives the system affects instruction in a positive way. Therefore, in designing an assessment framework for college students' critical thinking, it would be remiss not to envision the potential impact of such a framework on instructional design and practice.

What is the impact of the framework proposed in this paper? Does it provide a good foundation for a test worth teaching to? This question really has two parts. First is the question of the theoretical strength of the framework: Is the framework theoretically substantive and rich? Does it provide a powerful conception of good thinking, one that makes sense to impart to college students? We submit that the answer is yes. Most of this paper can be seen as an argument for the theoretical soundness of the framework. In particular, the intent of the rationale section preceding this section was to lay out the full answer to the question of theoretical strength: It provided a justification for each aspect of the framework and showed why each was central and important to good thinking.

The second part of the question of whether this framework yields a test worth teaching to concerns practicality. Does the framework provide accessible elements around which to organize instruction? This question is crucial, for it invites us to look into the future. It invites us to envision the potential impact of the proposed framework on college-based instruction. So what might it be like to "teach to the test" when the test in question measures a set of key thinking dispositions and abilities in terms of how they play out in a few basic problem types?

Such a scenario is not difficult to envision. To begin with, instruction would target the techniques, strategies, and skills associated with each broad disposition and ability. For example, cultivating the disposition and ability to think broadly would involve teaching students to do such things as probe assumptions, actively seek multiple points of view, brainstorm, and think flexibly. Cultivating the disposition and ability to be metacognitive would involve teaching students how to track and evaluate their own thinking and to strategically work through thinking challenges.

But just teaching these competencies is not enough. The motivation to use them must be built into the instruction-assessment cycle. So, for example, physics instruction and tests might evaluate students on how broadly they think about ways to invent an experiment, and how metacognitively and strategically they work through a problem of explanation in physics.

Notice that basic problem types come into play quite naturally. For instance, as just related, physics students can be asked to think broadly about designing an experiment: *Invention & design* is a basic problem type. They can be asked to be metacognitive about constructing an explanation for a physics phenomenon: *Explanation,* too, is a basic problem type. This observation highlights an important instructional implication of the assessment framework presented: ***Subject matter instruction that cultivates good thinking dispositions and abilities can easily be organized around basic problem types.***

Here is an example. Imagine a college course in comparative literature in which students are studying 19th century fiction. The basic problem types of explanation and justification might be used as a foundation for lessons that teach students to examine, say, the comparative effects of the industrial revolution on British and French novels. For example, by using the standards inherent in the problem type of explanation, students can learn to do such things as construct interpretations using supporting evidence. Learning to do this well involves being intellectually careful (a disposition and ability), being metacognitive and strategic (another disposition and ability), and other related dispositions and abilities. In subsequent lessons that are organized around the basic problem type of justification, students can learn to construct and support views concerning the relative merits of different 19th century authors and genres, using standards related to the use of evidence and attention to multiple viewpoints. Again, the appropriate dispositions and abilities will come into play as students work through the problem type.

It is hard not to agree on the value of a comparative literature course such as the one sketched above. If taught competently, such a course would contribute to the development of students' critical thinking while at the same time enriching their understanding of the subject matter. Moreover, not only comparative literature but virtually all college subject matters would gain by being organized in terms of basic problem types. This is no accident: the main reason for using the conception of basic problem types in an assessment framework is precisely that it characterizes the general types of thinking challenges that are at the heart of subject matters and that typically cross domains and contexts. For example, justification—a basic problem type—is important in science and equally important in comparative literature: Decision making—a basic problem type—is as important in political science as it is in anthropology. This same point can be made for the remainder of the basic problem types.

This paper is not the place to work out detailed recommendations for college curricula and instructional design. The important thing is that the proposed assessment framework suggests that such recommendations are possible, and indeed promising.

Tests *do* drive education, and they will most likely continue to do so. Let us take this as an enjoinder to adopt an assessment framework that sets a high and gracious standard for what instruction in a college setting might be and mean.

References

Baron, J. (1985). *Rationality and intelligence.* New York: Cambridge University Press.

——— (1991). Beliefs about thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal Reasoning and Education.* Hillsdale, NJ: Erlbaum.

Beardsley, M. C. (1975). *Thinking straight: Principles of reasoning for readers and writers.* Englewood Cliffs, NJ: Prentice-Hall.

Black, M. (1952). *Critical thinking: An introduction to logic and scientific method.* Englewood Cliffs, NJ: Prentice-Hall.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-75). Hillsdale, NJ: Erlbaum.

Collins, A. & Ferguson, W. (in press). Epistemic Forms and Epistemic Games. *Educational Psychologist.*

CSTEEP. (1992). *The influence of testing on teaching math and science in grades 4-12.* Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Dewey, J. (1922). *Human Nature and Conduct.* New York: Holt

Ennis, R. H. (1969). *Logic in teaching.* Englewood Cliffs, NJ: Prentice-Hall.

——— (1986). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. S. Sternberg (Eds.). *Teaching thinking skills: Theory and practice* (pp. 9-26). New York: W. H. Freeman.

——— (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18* (3), 4-10.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis.* Cambridge, MA: The MIT Press.

Ericsson, K. A. & Smith, J. (Eds.) (1991). *Toward a general theory of expertise: Prospects and limits.* Cambridge: Cambridge University Press.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction, Executive summary.* Millbrae, CA: California Academic Press.

Gentile, C. (1992). *Exploring new methods for collecting students' school-based writing: NAEP's 1990 portfolio study.* Washington, DC: National Center for Education Statistics.

Gifford, B. R. & O'Connor, M. C. (Eds.). (1991). *Changing assessments: Alternative views of aptitude, achievement and instruction.* Norwood, MA: Kluwer Publishers.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39*, 93-104.

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science 16*, 395-429.

Greenwood, A. (1992). *National assessment of college student learning: Getting started.* Washington DC: National Center for Education Statistics.

Kuhn, D. (1991). *The Skills of Argument.* Cambridge, England: Cambridge University Press.

Loacker, G. (1991, November). *Designing a national assessment system: Alverno's institutional perspective.* Paper prepared for the study design workshop National Assessment of College Student Learning: Issues and Concerns, Washington, DC.

McPeck, J. E. (1981). *Critical Thinking and Education.* New York: St. Martin's.

Mentkowski, M. (1991, November). *Designing a national assessment system: Assessing abilities that connect education and work.* Paper prepared for the study design workshop National Assessment of College Student Learning: Issues and Concerns, Washington, DC.

Newell, A., & Simon, H. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental process. *Psychological Review, 84*, 231-259.

Norris, S. P. (1991). Informal reasoning assessment: Using Verbal Reports of Thinking to improve multiple-choice test validity. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal Reasoning and Education.* Hillsdale, NJ: Erlbaum.

Norris, S. P. & Ennis, R. H. (1989). *Evaluating Critical Thinking.* California: Midwest.

Nummedal, S. G. (1991, November). *Designing a process to assess higher order thinking skills in college graduates: Issues of concern.* Paper prepared for the study design workshop National Assessment of College Student Learning: Issues and Concerns, Washington, DC.

Ohlsson, S. (in press). Abstract Schemas. *Educational Psychologist.*

Passmore, J. (1967). On teaching to be critical. In R. S. Peters (Ed.), *The concept of education.*

Paul, R. (1986). Dialogical thinking: Critical thought essential to the acquisition of rational knowledge and passions. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 127-148). New York: W. H. Freeman.

———. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world.* Rohnert Park, CA: Center for Critical Thinking and Moral Critique, Sonoma State University.

Paul, R. W., & Nosich, G. M. (1991, November). *A proposal for the national assessment of higher-order thinking at the community college, college, and university levels.* Paper prepared for the study design workshop National Assessment of College Student Learning: Issues and Concerns, Washington, DC.

Perkins, D. N. (1981). *The mind's best work.* Cambridge, MA: Harvard University Press.

———. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77*(5), 562-571.

———. (1986). *Knowledge as design.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

———. (1992a July). *The hidden order of open-ended thinking.* Invited paper presented at the Fifth International Conference on Thinking, Townsville, Australia, July 6-10, 1992.

———. (1992b). *Smart schools: From training memories to educating minds.* New York: The Free Press.

Perkins, D. N., & Salomon, G. (1987). Transfer and teaching thinking. In D. N. Perkins, J. Lochhead, & J. Bishop (Eds.), *Thinking: The second international conference* (pp. 285-303). Hillsdale, NJ: Erlbaum.

———. (1989). Are cognitive skills context bound? *Educational Researcher, 18*(1), 16-25.

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss, D. N. Perkins, and J. Segal (Eds.), *Informal reasoning* (pp. 83-105). Hillsdale, NJ: Erlbaum.

Perkins, D. N., Jay, E., & Tishman, S. (in press). Beyond abilities: A dispositional theory of thinking. *The Merrill-Palmer Quarterly.*

Perrone, V. (Ed.). (1991). *Expanding student assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Rabinowitz, M., & Glaser, R. (1985). Cognitive structure and process in highly competent performance. In F. D. Horowitz & M. O'brien (Eds.), *The gifted and talented: Developmental perspectives* (pp. 75-98). Washington, DC: American Psychological Association.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*(2), 113-142.

Scheffler, I. (1981). *The Anatomy of Inquiry.* Indianapolis: Hackett

Schwartz, J. L., & Viator, K. A. (Eds.). (1990). *The prices of secrecy: The social, intellectual, and psychological costs of current assessment practice.* Cambridge, MA: Harvard Graduate School of Education, Educational Technology Center.

Scriven, M. (1991a, November). *Edward Morante on the New Jersey Test.* Critique prepared for the study design workshop National Assessment of College Student Learning: Issues and Concerns, Washington, DC.

———. (1991b). *Multiple-rating items.* Point Reyes, CA: Pacific Graduate School of Psychology.

Simon, H. A. (1981). *The sciences of the artificial.* Cambridge, MA: The MIT Press.

Sternberg, R. J. (1985). *Beyond I.Q.: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Tishman, S., Jay, E., & Perkins, D. N. (in press). Thinking dispositions: From transmission to enculturation. *Theory Into Practice.*

Toulmin, S.E. (1958). *The uses of argument.* Cambridge, England: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1979). *An introduction to reasoning.* New York: Macmillan.

Appendix: Sample Assessments

The following sections use sample assessment situations to illustrate how the proposed two-part framework can be practically implemented. The key idea is to engage the test-taker initially in a relatively unstructured version of a basic problem-type task and later in a more structured version. Comparison of performance on the two allows the differentiation of dispositions and abilities. The first example looks at assessment in a common knowledge context, while the second example focuses on the specific content domain of physics.

Example 1: Sample Assessment in the Common Knowledge Context

The purpose of assessment in a common knowledge context is to measure how well college students are able to tackle important, everyday thinking challenges that do not depend on specialized knowledge—situations, for example, like deciding for whom to vote or what car to buy, solving household problems, and communicating with friends and colleagues. Below is such an assessment in the basic problem type of decision making. It involves two tasks. Task one is designed to assess overall dispositions and abilities. Task two, when preceded by task one, is designed to measure abilities and distinguish them from dispositions. In a real test situation, the subject would not be exposed to task two until after having completed task one.

Task One

In order to reduce health problems related to smoking, your state is considering a new 30 cents tax on cigarettes. Some say that such a tax would deter many people from smoking. Others say that it is unfair to place yet another tax burden on citizens. There may be other options. Suppose you are in a position to make a recommendation to policy-makers. How would you think through this issue, and what would you decide to recommend?

Show the process of your thinking on paper as you work through the problem.

Task Two

Here is a problem in three stages. At each stage, show your thinking process as fully as possible on paper.

Stage one: Look back at the cigarette-tax problem you worked on earlier. You may have identified a few options before deciding which one was best. Now, see if you can brainstorm several options (including those you mentioned already), thinking broadly about possible alternatives.

Stage two: Select the two options that you paid the most attention to in Task One and seek reasons for and against them. Try to seek a variety of reasons, reflecting several perspectives.

Stage three: Choose which option you think is best, and justify your choice.

Remember, try to show the process of your thinking as best you can

A Sample Response to Task One

A college senior, we'll call him "John," is given task one (he has not yet seen task two). He reads over the test question and immediately favors the "no tax" option.

John thinks, "People should be able to smoke without having to pay an arm and a leg for it. After all, if folks want to kill themselves, it's their business."

Yet, John can see the other side of the case, too. "Everyone knows the dangers of smoking," he thinks. "It's really a shame that young kids are still taking up the habit."

John starts recording his thinking:

My first leaning is toward the "no tax" idea, because it seems unfair to taxpayers. But I can appreciate the other side, too. Tobacco-related illnesses must cost the State plenty. Not everyone has insurance, and someone has to pick up the medical costs. Which idea is best, tax or no tax? Reading over the problem, I notice there is nothing saying that these are the only two possible alternatives. Maybe there are others. I should make a list.

1. *Tax cigarettes; use the revenues to pay for health costs related to smoking.*
2. *Don't tax cigarettes: let people do what they want.*
3. *Give a tax credit to people who don't smoke!*

BEST COPY AVAILABLE

Intrigued by the tax credit option, John stops brainstorming. The idea of a tax credit sounds great: it seems to solve the problem in a creative way. Nonetheless, John recognizes that it is important to think carefully about this option—there may be some hidden difficulties with it.

"What are the pros and cons?" he asks himself. On paper, he records his reasoning like this:

Option: Give a tax credit to people who don't smoke

Pro	**Con**
• *It helps people who don't smoke, but doesn't create more of a tax burden for those who do.* • *People like to make extra money: this will be an incentive for people to quit smoking.*	• *It won't create extra revenues for the State— revenues that may be needed to pay for smoking-related medical costs already incurred.* • *It would be hard to enforce: How could you tell if people were telling the truth when they claimed they didn't smoke? You'd have to have a doctor's certificate or something.* • *It probably can't be a very big tax credit, so maybe people won't see it as worthwhile.* • *Many people start smoking when they are young, before they have begun to pay taxes themselves. So it may not deter people from starting smoking initially.* • *It could be administratively costly to implement.*

John regrets that there are so many drawbacks to the tax credit idea—it had seemed so promising! He writes:

Initially I was enthusiastic about the idea of tax credits. But I see that it has many negative aspects. I think the best thing to do is to leave the situation alone: don't increase the tax on cigarettes. Although a tax may deter some people from smoking, the overall economic burden it places on the taxpayer is unfair. And, in these economic times, it will be felt particularly *harshly. My recommendation to policymakers would be to not impose any new taxes, and to look for other ways to raise money to pay for cigarette-related medical costs already incurred.*

Scoring Task One

Which dispositions and abilities should serve as standards to assess John's performance? Recall that there are seven broad disposition and ability categories. Not all basic problem types necessarily call into play all seven in a central way. For instance, the disposition towards sustained intellectual curiosity seems not to play a key role in the problem type of decision making in general, although it might in some specific cases. How can we decide which dispositions and abilities to target?

The answer is found by looking closely at the problem type in question. For each basic problem type there is a cluster of dispositions and abilities that are typically key to good thinking in that problem type. For example, the disposition and ability to be broad and adventurous is key to good decision making, because it underwrites the search for innovative ideas. And the disposition and ability to seek reasons and evidence is key, because it mitigates against impulsiveness and justifies intended action.

Part of the answer, then, is to target the dispositions and abilities central to the problem type. The other part of the answer is to take into consideration the exigencies of the task at hand—for example, whether, in a particular decision situation, there are only one or two viable options, or a limitation on available resources. Thus, some decision-making tasks will invoke a full cluster of dispositions, other special cases may invoke only one or two.

Each disposition and ability within a cluster can be broken down into a set of yardsticks. These are scales that rate levels of proficiency. They identify specific thinking behaviors to look for, and measure subjects' strengths and weaknesses therein. For example, some of the yardsticks that measure "broad and adventurous thinking" might include the following:

Broad and Adventurous

- Generates several options/ideas
- Generates diverse options/ideas
- Explores alternative perspectives/points of view

Performance on each yardstick can be rated on a scale of 0-4, in which 0 stands for "not present" and 4 stands for "richly present." The numerical rating levels would be established by devising prototypical responses for each level. For instance, piloting might show that generating eight options is a high degree of brainstorming for a given student grade or age level. In such case, a level 4 rating on the "generates several options" dimension would be given if a student generates eight or more options, a level 0 would be given if no options were generated, with levels 1-3 given for responses in between. Meanwhile, a level 4 rating on the "generates diverse options" dimension might be given to performances that diverge in three or more ways from the obvious options given in the problem statement.

The cigarette tax problem certainly calls for broad and adventurous thinking, both in the initial search for alternative options, and in the subsequent search for reasons for and against those options. John's performance is thus easily gauged. For example, he generates three options—not a large quantity, but at least minimally beyond the given. By comparison with prototypical response levels, let us suppose his score on this yardstick would be a "1." However, one option—the idea of a tax credit—does dramatically diverge from the obvious options. This rates a "2" on the diversity dimension.

What about the "explores alternative points of view" yardstick? Looking at John's pro/con list of reasons concerning the tax credit, we see that he does indeed examine some different perspectives. For instance, in addition to looking at the situation from the perspective of the taxpayer, he considers the perspectives of nontaxpayers, medical personnel, and administrators. John is strong here: comparison with the prototypical responses will tell us precisely how high a score in this yardstick his performance warrants.

As mentioned earlier, broad and adventurous thinking is not the only disposition and ability that is relevant to decision making. Seeking reasons and evidence is also key. This is because good decision making involves identifying balanced and well-justified reasons underlying possible options. Some of the yardsticks that measure "reason-seeking behavior" include:

Seeking Reasons and Evidence

- Seeks both pro and con reasons
- Seeks evidence to support reasons
- Connects reasons to context

How does John's performance measure up? Well, he does a good job of seeking pro and con reasons when he explores the idea of a tax credit. But later, when he returns to the "no-tax" option, he fails to seek unfavorable reasons, and only identifies reasons in support of this alternative. Nonetheless, all of the reasons he articulates do connect logically to context: that is, all of them are relevant to the issue at hand. Again, comparison with a prototypical response can indicate exactly what values to assign to these specific behaviors.

Compiling scores. The 0-4 average of all the yardstick scores together indicates overall dispositions and abilities performance (a good decision maker should rate highly in all key disposition and ability categories). The sum of scores within each category indicates the strength and role of that particular disposition and ability within the performance (John, for example, would seem to warrant an overall higher rating for the disposition and ability to seek and evaluate reasons than he would for the disposition to be broad and adventurous). The distribution of scores within a disposition and ability category discriminates more specific strengths and weaknesses (identifying, for example, a strength in generating several options, but a weakness in probing assumptions).

Task Two: Distinguishing Between Dispositions and Abilities

Earlier it was mentioned that a *sequence* of performance tasks is necessary to distinguish between dispositions and abilities. This distinction is important, because effective thinking in everyday contexts tends to be dispositions driven. Most thinking challenges that people face are broad and unstructured. In decision making, for example, the decision point is usually presented as a dichotomy ("decide whether to study or go to the movies"), and one is not told to think broadly or to reason carefully—one is merely asked to make a choice. This is more or less the scenario in version one of the cigarette tax problem. John is not told what specific steps to take to think through the problem; he is simply asked to think it through and show his work.

The unstructured task format simulates how most real-life thinking challenges are presented, and assessing subjects' thinking on unstructured tasks provides an overall dispositions and abilities profile. But it doesn't yet discriminate *between* dispositions and abilities. So far, John's spontaneous performance indicates what he is disposed and not disposed to do *combined* with what he is able to do. For instance, we know that he is not disposed to generate multiple options, simply because he did not. But we do not yet know whether he lacks the inclination to do so, or the ability.

To isolate John's decision-making abilities, it is necessary to assess his performance on a structured task following his performance on an unstructured one. Recall that task two of the cigarette tax sequence *is* structured. It asks John to revisit his earlier work and expand it according to specific stages of decision making. For instance, it explicitly asks him first to brainstorm options, and then to seek pro and con reasons. Structured thus, the explicit task requirements "stand in" for the dispositional elements. John does not need to be *disposed* to think broadly. The problem statement directly tells him to do so.

A Sample Response to Task Two

Here is a sample of John's performance on task two of the cigarette tax sequence. Of course there would be other parts of his response, but this much serves to make the point. Starting at stage one, where the task asks him to think broadly about several options, John records his thinking like this:

Stage One: Brainstorm Options

1. *Tax cigarettes, use revenues to pay for health programs related to tobacco use.*
2. *Tax cigarettes, use revenues to educate people about the dangers of smoking.*
3. *Tax cigarettes, use revenues to assist families who have had a parent die from a tobacco-related illness.*
4. *Give a tax-credit to people who don't smoke.*
5. *Impose a tax only on parents of minors who smoke, so they will exert more influence over their kids.*
6. *Don't tax anyone, but don't help them with smoking-related health problems either.*
7. *Tax only people under 20, to dissuade young people from starting smoking.*

Scoring Task Two

Precisely the same yardsticks used in task one are appropriate here, except that this time the overall score will indicate abilities only, not dispositions-plus-abilities. For example, recall that two yardsticks for broad and adventurous thinking are: "generate several options" and "generate diverse options." Clearly, John has the ability to generate many more options than he was inclined to do on the unstructured task. Earlier he identified three options; now he brainstorms seven.

On the other hand, although the quantity of John's options has increased significantly, their diversity has not. The tax-credit option is still the only idea that breaks set: all the rest stay within the original parameters of the problem—tax or no tax.

Comparing John's task one and task two performances indicates where his decision-making strengths and weaknesses are, at least for this task. If his performance on other decision-making tasks showed the same pattern, we could conclude that John needs to learn to see the appropriateness of generating many options at decision points. His task two performance shows that he is capable of this: his failure to do so in task one is a dispositional shortfall. John also needs to learn tactics for "breaking set"—for broadening his search to include diverse and unusual ideas. The fact that he failed to do so even when asked indicates (if corroborated) a shortfall in ability.

Levels of Proficiency

There is of course a more global question at stake. What information can data from task sequences such as the cigarette tax problem yield about subjects' overall levels of proficiency?

Or, put more simply, how can an analysis of such data reveal precisely how well college students think?

Recall that students will be tested in different problem types, and in different contexts and domains. So, for example, a battery of tests might include tasks in the problem types of decision making, explanation, and problem solving. These tasks will occur in at least two contexts: common knowledge and the student's academic major. The data obtained from such testing lend themselves to the following kinds of analyses.

To begin with, a single numerical index might be assigned to each dispositions and abilities performance in a task, then averaged across problem types within a domain to yield an aggregate dispositions and abilities score in that area. For example, a subject may show a common knowledge profile that is strong in broad and adventurous thinking but weak in strategic planning. The same can be done within problem types within a domain to indicate levels of proficiency within problem types. For instance, scores might be aggregated to indicate how good of a problem solver, overall, a subject is within a domain of history.

Also, a analysis can look for correlations of the same dispositions and abilities across domains, indicating whether (and which) dispositions and abilities tend to be general or domain specific. When substantial positive correlations obtain, numerical indices across domains can be averaged to yield a measure of how proficient a thinker is for that disposition and ability. Also, one can look for correlations among dispositions and abilities. For example, one might find that the disposition and ability to be metacognitive correlates highly with the disposition and ability to think strategically, but not so highly with the disposition and ability to be intellectually curious.

Of course, care must be taken not to oversimplify. For example, it is important to look closely at each kind of aggregation to determine the degree of variation among pooled scores and whether such scores reflect anything more than an average of very diverse individual scores. Other caveats apply as well. Nonetheless, there is good reason to hope that with appropriate care, analyses of data from tasks like the cigarette tax problem can provide a rich and reliable picture of dispositions and abilities levels of proficiency in college students' thinking.

Example 2: Sample Assessment in a Content Domain

In this brief example of an assessment task in the content domain of physics, the basic problem type being assessed is that of constructing an explanation. In science, an important part of explanation building involves the generation and testing of hypotheses.

The Assessment Task

The following is an unstructured assessment task in physics that would be given as Task One, similar to the previous example. Normally, a structured Task Two that distinguishes between dispositions and abilities as in the previous example would be given as well, but the Task two version is not presented here for the sake of brevity.

Task One

A scientist is sitting in a fixed-position seat in the center of a windowless cabin in the space shuttle traveling in outer space. Since she is completely isolated inside the cabin away from the pilot's cockpit, she does not know anything about how the space shuttle is moving. In general, the scientist feels weightless, but when she lets go of a pencil an arm's length in front of her, she notices that it drifts slowly off to her right and hits the right wall of the cabin. She releases another pencil and it drifts to and lands on the left wall. She repeats these actions two or three times with exactly the same results. How might the scientist explain this?

Try to generate hypotheses that will help you figure out what is going on and how to test them. Record your thinking and reasons. Be specific about details. Qualitative responses are sufficient, although you may provide equations or quantitative support if you wish. Use diagrams, if helpful. Remember that the aim is to construct the best investigation and explanation you can about the movement of the space shuttle.

A Sample Response

Margaret is a first-year college student who has had high school physics. She reads the problem and is first concerned about figuring out what kind of movement is involved, for example, linear, curved, spinning, zigzag, and so on. So she begins to think on paper as follows:

> *I'm not sure but I think the space shuttle is not moving in a straight line because offhand it seems to me that the pencil would always land in the same place if it were moving in a straight line. Anyway, so that is my first hypothesis I want to test.*
>
> *HYPOTHESIS: The space shuttle is not moving in a linear motion.*
>
> *If the space shuttle were moving in a straight line, what would happen? I know from Newton's Laws that a body in motion will stay in motion unless there is a force upon it. So, if the space shuttle, the scientist, and the pencil were all moving at some velocity, when the scientist releases the pencil it should still travel at the same velocity. But the scientist and the space shuttle are also still going at that same velocity, so I guess to her it would look like the pencil isn't moving at all. It would be suspended in mid-air. It wouldn't land on the wall at all. Anyway, in the problem, the scientist sees the pencil*

move to the wall, it does not stay still, so the space shuttle is not moving in a linear motion.

There's one problem though. The pencil might not look like it's standing still if the space shuttle is accelerating. I have to make sure it is not moving in a linear motion but also accelerating or decelerating.

> *HYPOTHESIS: The space shuttle is not accelerating or decelerating in a linear motion.*

If it was accelerating, there would be a force on the space shuttle, the scientist, and the pencil, according to the formula F=ma. But when the scientist releases the pencil, there is no longer a force on it, then there is no acceleration. F = 0, so a = 0. The pencil stops accelerating, but the space shuttle and scientist are still accelerating, so the space shuttle would "bump into" the pencil. That means, to the scientist, it would look like the pencil is moving straight backwards (or forwards, if the space shuttle was decelerating). That's not what she sees in the problem, so it's not accelerating or decelerating in a straight line.

> *HYPOTHESIS: The space shuttle is moving in a spinning motion.*

If it was spinning, there is a force on the space shuttle, the scientist, and the pencil. Like in the linear acceleration example I did, when the pencil is released, it stops accelerating. But the space shuttle and the scientist keep going around. So to the scientist, it looks like the pencil goes flying out toward the edge of the space shuttle. I think it's centrifugal force, like when you whirl around a ball on the end of string and then let it go. It flies off in the direction it was going when released, so it makes sense that if you release it at different places it will land on different spots on the edge. I think this hypothesis is right: the space shuttle is spinning. And I think it is spinning to the left because the pencil appears to go to the right.

Believing that she has figured out the pattern of movement, Margaret goes on to determine other aspects, like the axis of rotation, in the following way:

I thought of a great way of showing that it is spinning.

> *HYPOTHESIS: If the space shuttle is spinning and an object is released directly on the axis of rotation, it will not appear to move.*

From most starting points, the pencil will move to the edge, BUT if you release right on the rotating axis, it won't move. If the scientist does this experiment, she can prove that the space shuttle is spinning.

Margaret concludes with the following final explanation:

> *I think that the space shuttle is spinning counterclockwise in space because it is consistent with the evidence. I have eliminated several other possible kinds of movement.*

Scoring the Task

Several disposition and ability categories are relevant to the explanation and the subproblem type of generating and testing hypotheses. For purposes of illustration, two can be examined: the disposition and ability to build knowledge and understandings, and the disposition and ability to seek reasons and evidence. Potential yardsticks for assessment include the following:

Building knowledge and understandings

- Considers and eliminates competing hypotheses
- Connects to and builds on prior knowledge and findings
- Identifies and justifies, and makes explicit operating assumptions to proceed

Ratings of performance for these dimensions would again use 0-4 rating scales that are based on prototype responses. For example, on the "considers and eliminates competing hypotheses" dimension, a low score would be given if a subject considers only one hypothesis or only non-competing hypotheses, while a high score would mean that a subject considers a number of competing hypotheses, eliminating those disconfirmed by the evidence. Similar scales would be constructed for the other yardsticks.

Margaret's performance can be gauged according to the above yardsticks. In the course of her performance, she does not merely focus on one tack. Rather she entertains and eliminates several hypotheses before arriving at her final conclusion. She would score well on the "considers and eliminates competing hypotheses" dimension. In the dimension "connects to and builds on prior knowledge and findings," Margaret's hypotheses and tests do build on one another in a logical way. She uses findings from previous steps to construct her next hypothesis. Again, she would score well here. Regarding identifying ambiguities, the task presents a number of them, all of which go unrecognized. For instance, does "drifts to the right" mean that the pencil stays in the same horizontal plane? Margaret scores poorly on identifying ambiguities.

The disposition and ability to seek and evaluate reasons is especially central to hypothesis generation and testing because it is necessary to support claims and explanations using appropriate evidence. The following yardsticks could be employed to gauge performance:

Seeking reasons and evidence

- Seeks evidence to support claims and explanations
- Uses appropriate evidence
- Makes predictive use of evidence

As before, scoring could employ 0-4 rating scales established by prototype responses.

Margaret uses evidence persistently in her investigation. She seeks and employs evidence for every claim she makes. She uses both qualitative, logical reasoning and some theoretical substantiation (using formulas to support the reasoning). Across the whole thinking episode, she does a good job of mastering different kinds of evidence to arrive at her final explanation. Some evidence serves to eliminate possibilities, like linear motion; other evidence supports a theory, like the reasons in support of the spinning hypothesis; still other reasons are used as confirming evidence, like the axis of rotation proof. Margaret would score fairly well on both "seeking evidence" and "appropriate use of evidence."

Predictive use of evidence means knowing what qualifies as evidence before performing the experiment and projecting what different outcomes would mean for the theory in question. Margaret would score lower here. She mentions the prediction about releasing the center. But there are several other predictions she could make. For instance, if her hypothesis about spinning is correct the pencil should seem to follow a curved path as it moved toward the wall. But Margaret does not demonstrate in her text that she knows in advance what constitutes evidence for her hypotheses. While she may be able to do this when prompted, her score here would be on the lower side because it is not demonstrated in her spontaneous performance.

This example demonstrates an unstructured task that aims to capture the overall dispositions and abilities in question. As with the decision-making example outlined above, a second more structured task could be designed to separate out the abilities component. The level of proficiency guidelines outlined for the decision-making example, and the way in which the six recommendations are met by these assessment measures, apply to this example as well.

Central and Incidental Features of These Examples

The examples presented in the last two sections have the virtue of concretizing the five principles for an assessment framework proposed in the main body of the paper. They illustrate how one might translate the principles into actual assessments. However, such examples present a hazard as well. Too easily they can be taken as advocating more than they do. They may be read as votes for features that are in fact incidental to the approach developed here.

In general, the proposed plan leaves a number of matters open, even as it adopts specific stands on others. It may be useful to comment specifically on what features are central to the approach and what others are less so.

***Central:** Overarching dispositions and abilities and basic problem types much like those listed.* As mentioned earlier, the suggested abilities and dispositions and basic problem types of thinking reflect a collation of sources in the literature. They also have some modest claim to psychological reality: each disposition and ability and basic problem type can be viewed as a knowledge system involving a body of concepts, values, and strategies. Finally, all the dispositions and abilities and basic problem types make intuitive sense and have high relevance in academic, workplace, and citizenship contexts. Features such as these recommend something like the proposed scheme.

Incidental: The exact dispositions and abilities and basic problem types. The proposed dispositions and abilities and basic problem types represent an effort to achieve a reasonable framework, but items might be added to either list, or some entries split or consolidated.

***Central:** Authentic assessments that probe the thinking process.* Principles #3 and #4 argued for these features. In both cases, the rationale was simply one of genuine sampling: to measure thinking, one should draw something close to an authentic sample of thinking. Well-designed multiple-choice instruments are certainly suitable for gauging some aspects of thinking. But they cannot readily reveal features of the thinking process such as breadth of search nor can they easily differentiate dispositions and abilities.

Incidental: Written problems, essayistic responses, thinking on paper, exclusively authentic assessments. The examples offered in this paper adopt the format of written responses where students are asked to "think on paper." This is one possible approach with a number of tradeoffs. For example, easy to administer, it certainly does not probe thinking as deeply as gathering think-aloud oral protocols. For another example, formats using videotape or videodisc, while more costly, might well prove more motivating. Scriven (1991a) notes that responses on critical thinking instruments can be contaminated by essay-writing skills that have little to do with the thinking per se, a point of caution.

Authentic assessments need not be the only source of data. As discussed earlier, multiple-choice and multiple-rating items can probe certain thinking skills in a very precise way (e.g., Norris & Ennis 1989; Norris 1991; Paul & Nosich 1991; Scriven 1991b). It was argued that this kind of testing does not capture realistic episodes of thinking in academic, workplace, and citizenship contexts well enough to be the mainstay. But such techniques certainly could supplement authentic assessment. Moreover, the structured "task two" phase of the approach to testing in the examples given might well use multiple-choice and especially multiple-rating items in certain places.

Central: Rich tasks that require some time for completion. This is an efficiency issue. Authentic assessments that pose rich tasks and probe the thinking process necessarily take some time to complete. Such tasks commonly would take half an hour and could easily take an hour or two. High time per item is an inevitable cost of assessment that deals with authentic samples of thinking.

Incidental: High total test time per person. A seeming implication of the above is high test time per person. For instance, if each item takes an hour and one wanted to profile a person's thinking in several basic problem types and in both the common knowledge context and areas of specialization, the testing might take days. However, this much testing of an individual is not really necessary. Sampling techniques allow much less time per person, providing the aim is not a full profile of each test-taker. If, for example, the aim is to assess the overall impact of a program, one can sample across tasks and students in a way that dramatically reduces the time per person. Perhaps, 3-4 hours per student might do. Yet a different answer is that the process can be streamlined once more data are in hand. Initially, more information is needed to determine patterns of correlation across different disposition and ability clusters and basic problem types. Depending on those correlations, a much briefer assessment may prove possible.

Central: Differentiating dispositions and abilities. As argued in an early section of the paper, dispositions play a crucial role in determining how people think in realistic settings. Arguably, dispositional shortfalls are actually more problematic than shortfalls in know-how.

Incidental: Unstructured and structured passes at the same task. The examples addressed the problem of discriminating dispositions and abilities through a two-phase task structure, featuring an unstructured phase followed by a structured phase in which the test-taker responds to point-blank requests for various kinds of thinking about the same problem situation. Other designs certainly are possible. For instance, the assessment might involve an unstructured problem and a different structured problem. An instrument to differentiate sensitivities and inclinations would require another organization yet. The format used in the examples simply illustrates one reasonable possibility.

Relatedly, the test-takers may or may not know what the dispositions and abilities of interest are. Many factors recommend public standards, at least at some level of generality. It should be noted that public standards create a dilemma for the assessment of dispositions. Students who know the standards very likely would try to feign the dispositions.

There are several possible answers to this dilemma. One is that the public version of the standards would be rather general. It might in fact not help test-takers that much with the particular test tasks unless they had genuinely internalized the dispositions. Another is that a portfolio approach rather than a sit-down test might be employed, as in the recent portfolio assessment of writing by NAEP (Gentile 1992). It seems unlikely that students could feign target dispositions successfully across diverse tasks over months. Or, if they tried, quite likely they would tend to internalize the dispositions. No doubt there are other approaches as well.

Central: Rating of complex performances on multiple dimensions. If authentic assessments involving complex responses are involved, there is little choice but to rely on scorers' thoughtful ratings to render this raw data into a form that can be subjected to further processing—extraction of averages, examination of correlations, and so on. In the examples, scoring was based on "yardsticks," five-point scales that gauged how a particular disposition and ability applied in the context of a specific task involving a basic thinking problem type.

Incidental: five-point scales, specific ways of aggregating information. Rating scales pose a number of challenges that can be handled in various ways. One, for instance, is scale definition. The extreme ends of the scales can be anchored by various techniques and the levels in between defined in different styles. Different views of the appropriateness of averaging and other statistical techniques can be applied to the resulting ordinal scales. A good many technical problems invite attention here.

ASSESSING SPEAKING AND LISTENING: PRELIMINARY CONSIDERATIONS FOR A NATIONAL ASSESSMENT

by John A. Daly

Perhaps the single most important skill people possess is that of communication. In so many ways, the ability to speak and listen with, and to, others, defines us as human, that most sociable of species. We accomplish tasks, develop relationships, learn our most important lessons, and come to understand ourselves primarily through communication. What we say, and don't say, has incredibly important consequences for ourselves and others as does what is said, or not said to us. What is intriguing, then, is how little systematic work on the assessment of speaking and listening effectiveness has been conducted. Perhaps the paucity of assessment research on communication skills is because of the presumption that speaking and listening are almost too basic. Everyone talks and people appear to listen, so what's the big deal? But that assumption represents a crucial mistake. For what we know about communication belies the presumption: Many people in our nation are simply not that effective at even the most basic of speaking and listening skills.

The Importance of Communication

For many years now, scholars as well individuals in various organizational roles have emphasized the importance of communication to work effectiveness and life satisfaction. Study after study has asked business leaders, for instance, what they most wish they could see in college graduates. Without exception, leaders want new employees to have better communication skills than they currently have. That wish isn't surprising. Studies of what contributes to the success of individuals in work situations reveals the central role of communication. Curtis, Winsor, and Stephens (1989), for instance, asked more than 400 personnel managers in diverse industry groups to indicate the relative importance of various skills in obtaining a job and successfully performing in that job. In addition, these experienced personnel specialists were asked to identify ideal management profiles. In all three exercises, oral communication skills were at the top (Appendix A reproduces their results).

Findings such as Curtis, et al. (1989) are not surprising. People spend an enormous amount of their time communicating. Towards the beginning of the century, Rankin (1928) observed that people spend approximately 75 percent of their waking hours in some sort of communication. Of that time, 46 percent was in listening, 30 percent in speaking, 16 percent in reading, and 9 percent in writing. Years later, Klemmer and Synder (1972) found that in the workplace people devote about a third of their time conversing and in meetings. They also found that people consistently underestimate the amount of time they spend speaking and consistently overestimate the amount of time they think they spend writing. Luthens, Hodgetts, and Rosenkrantz (1988) found that effective and successful managers spend a large portion of their time speaking and listening. And, when it comes to listening, research suggests that employees of major corporations report spending more than half their time listening (Wolvin 1991).

The central place of communication skills to workplace success is not surprising since the relative amounts of time people spend speaking and listening in jobs appear to increase in line with work responsibilities. Top executives spend the vast majority of their time speaking and listening. Effective management, as well as successful leadership, are inextricably tied to good communication skills (Alexander, Penley & Jernigan 1992; Luthens, et al. 1988; Sypher & Zorn 1986). And, it isn't only that people in high-level positions need good communication skills, it is also the case that to get into those high-level positions, good communication skills are required (Curtis, Winsor & Stephens 1989). For in virtually any profession, success is tied to communication. In sales, the difference between having a product and selling it lies in communication. In education, the difference between knowing something and teaching it lies in communication. Communication is related, as well, to organizational productivity (Papa & Graham 1991).

In nonwork situations, communication is just as key. Relationships are communication events (imagine, for instance, a relationship without communication). Marital satisfaction is tied to communication (Gottman 1979; Huston & Vangelisti 1991; Noller & Fitzpatrick 1990). Bettering couples' communication skills at, for instance, handling disagreements can significantly reduce the likelihood of their marriage ending. Life expectancy is related to social support which, inherently, is created and maintained through communication (Duck and Silver 1990). Children are taught values, beliefs, and practical life skills through parental communication.

Given its importance, one would assume that people are generally competent at communication. While there is relatively little data on this issue, what does exist suggests that assumption may not be correct. Vangelisti and Daly (1989), for instance, using national survey data, discovered that upwards of 25 percent of young Americans (ages 18-24) could not perform some very basic communication tasks such as giving directions. Study after study suggests that in work situations employees sorely lack effective speaking and listening skills (DiSalvo 1980; Hanna 1978; Henry & Richmond 1982).

The case is straightforward: communication is critical to people's lives, personally and professionally, yet many people are not that effective at it.

Basic Conceptual Concerns

Creating assessments that tap people's competencies at speaking and listening offers numerous opportunities, politically, educationally, and theoretically. We need to grasp just how well people communicate. Politically, not knowing this continues a very vicious but hidden form of discrimination. For communication is how we accomplish most of life's chores and those who aren't skilled suffer the everyday insults of inefficiency and rejection. Educationally, for better or worse, assessments have a powerful impact on what is taught. They provide guidelines, shape new priorities, and offer systematic feedback. In most school systems in this nation, whether it be elementary or secondary, communication is not taught systematically. Most people first encounter structured communication courses at the college level. Assessments of college students, communication skills will, in the long run, filter down into elementary and secondary levels which in turn will encourage those institutions to more systematically teach students that most basic of skills--communication. Finally, assessment forces scholars of communication to carefully define their concepts, placing them within, if you will, "reality-based" contexts.

But assessments of communication skills face certain conceptual presumptions that ought to be noted as precursors to any discussion of assessment. Four come immediately to mind.

Communication is inherently interactive and dialogic. What I say has very little meaning separate from your response. Two individuals create meaning together in a social interaction, not individually; people, in a social exchange, mutually devise the interaction. Conceptually, then, the critical focus of attention needs to be on the "interact" rather than the "act." In social interactions such as conversations, interviews, negotiations, meetings, and small group exchanges, what is created is a joint function of the participants. A speaker and his or her listener sculpt meaning, revealing as they do their identities. They co-create and co-maintain a social reality that exists as long as they interact (and, indeed, in many ways, long after the actual interaction has ended.

Communication occurs in real time. There is no chance to revise in speaking. Once something is said, it can't be changed. Preparation for most sorts of interaction (aside from most public speaking and some very limited formal communication activities) occurs virtually simultaneously with production. It is unusually, in conversations and meetings to see people spend much time preparing prior to the interaction in terms of what they will say. Indeed, this is almost impossible since what you say will affect, immediately, what others say, and what others say can radically change what you are saying. This distinguishes speaking from writing. In writing, revision is anticipated. Forms of written communication are almost always time delayed. Conversation, on the other hand, requires immediate interaction between the parties involved.

Communication is inherently contextual. People speak and listen within contexts that shape the sorts of interactions that occur, the interpretations that arise, and the relative effectiveness of the communication. What is appropriate and effective in one context can easily be viewed as inappropriate and ineffective in another context. A formal speech such as what one might hear at a large annual meeting would be quite inappropriate at a small gathering of high school friends at a party. Similarly, something said to a person of higher status may have a completely different meaning than the same thing uttered to a person of lower status or power (Spitzberg & Brunner 1991). What is skillful in one context is not skillful in another. It is important the we, as a collection of scholars and interested parties, consider carefully whether a "national standard" is sensible.

Communication is personal as well as social. Far more than any other "basic" competency (i.e., reading, writing, computing) an individual's communication is their "own." We are, in many ways, our talk. Consequently, people need to carefully respect a person's or group's right to communicate as they deem efficient and appropriate. To attempt to homogenize talk, or to assume that there is one agreed-upon standard, in all contexts, is foolhardy. Readers of this paper are urged to carefully read Farr's response for more extensive discussions of the importance of insuring the personal nature of communication. There are important, interesting, and relevant differences among ethnic groups, income levels, genders, and education levels, to mention but a few. These differences, taken together, introduce an important constraint, or perhaps opportunity, in any assessment

Communication has important outcomes. As obvious as this may sound at first reading, it is critical to underline how central communication is to our lives. When people participate in a communicative interchange there are numerous and important consequences. People leave with a different sense of self, understanding better, perhaps, the others involved in the exchange. Beyond understanding, people accomplish their goals via communication. We teach, we sell, we learn, indeed we cope in our social world because of communication.

Issues in the Assessment of Communication Skills

Given these presumptions, the assessment of speaking and listening presents some tough challenges for any full-fledged assessment. Consider the challenges that will face any assessor:

Any assessment of communication must include a behavior sample. It is essential for validity purposes as well as practical political reasons to include, in any assessment of communication, a substantial behavior component. Knowing what to do is not the same as doing it. Communication happens aloud with others present and that must be assessed.

There is a need to carefully consider outcomes in addition to performances. Consider the following case: An individual delivers a very "effective" speech attempting to persuade an audience of some proposition. The speech is "effective" in terms of all of the standards one typically finds in textbooks on communication skills (e.g., interesting delivery, varied tone, good structure, effective evidence, etc.). However, when the audience is polled, while everyone feels the speech is a good one, technically, no one is persuaded. The speaker has done everything right but the outcome wasn't what he or she desired. Too often in assessments, we fail to consider that a really important consideration is whether or not the individual actually accomplished his or her interactive goal. To the extent this is true, any true assessment of communication performance needs to be listener centered rather than speaker centered. The critical thing is what the listener leaves the communication event with: does he or she leave understanding the message, is he or she persuaded by it, does she or he comes to like the communicator more, and so on. What the speaker does, separate from a listener, is unimportant.

One should note that some scholars, in studying communication effectiveness, argue that a focus on either knowledge or performance, per se, is not the appropriate emphasis. Spitzberg and Hurt (1987) are typical of this stream of thinking. They propose that:

> ...competence is an evaluative impression of behavior...the quality of a performance does not inhere in the performance itself, but in the evaluations of that performance. We cannot tell our students that if they establish eye contact, ask questions, manage their turn-taking, etc., they will be seen as competent. In any given context, they may enact these behaviors and still be viewed as entirely incompetent by the interactants involved. (p.29-30)

In other words, the best judgement of competence is in the evaluation of the other interactants. You are a competent performer if I, as your conversational partner, believe you are. On the face of it this seems quite reasonable. But questions emerge on closer analysis. What does one do when different partners have different standards? (The response is: that is what we call audience analysis--you need to adapt to different audience members.) But aren't there some standards we can all agree upon regardless of who is the listener or audience? (Sure, goes the response, but those standards are so elementary that we may not be assessing anything important.)

Assessing speaking competencies will be significantly more difficult, in practical terms, than assessing listening competencies, writing, or reading skills. Obtaining valid and reliable indices of an individual's capabilities in public speaking, meetings, and conversations will be a costly exercise--far more costly than assessments where paper and pencil suffice as measuring instruments. In the recent report (National Center for Education Statistics 1992) on the National Adult Literacy Survey, researchers at the ETS, along with a number of panels of experts, explicitly excluded assessments of "teamwork skills, interpersonal skills, and communication skills." The reasoning presented in the paper was that those three very important skills failed to fit their definition of literacy ("Using printed and written information to function in society, to achieve one's goals and to develop one's knowledge and potential." p. 11). Aside from the very obvious narrowness of the definition, the exclusion flies in the face of much scholarship on literacy (e.g., Ong 1982). But one might hypothesize that the exclusion of the most basic of human literacies--speaking and listening--was not conceptual at all. Rather, it may well have been for methodological and fiscal reasons. Assessing speaking and listening is much more time consuming, expensive, and conceptually complex than reading, writing, and computing.

One important consideration in any assessment is the role of nonverbal behaviors. While it is clearly more difficult to simultaneously examine both verbal and nonverbal behaviors as they relate to communicative accomplishments, it must be done. Why? Because people in social interactions, meetings, presentations, and any other sort of exchange use both their words and their actions to communicate (Knapp & Hall 1992). Moreover, interactants interpret what was said in the context of nonverbal behaviors. There are, luckily, some recent attempts to assess nonverbal communication skills (e.g., Rosenthal, Hall, DiMatteo, Rogers, & Archer 1979). However, as the assessment proceeds, scholars will need to more carefully identify nonverbal behaviors that contribute to effectiveness both in terms of encoding and decoding.

Most communication that counts happens among people who know one another to some degree. One of the interesting parameters of many communication encounters is that when we talk with someone we bring a history of interactions with that person to the exchange. We talk to friends, colleagues, spouses, and customers far more than we speak to strangers. This has consequence for any assessment. The question of whether people should give a presentation to a group of strangers or friends is an important one. How a person presents, what he or she assumes an audience knows, and even whether or not he or she will be judged competent depends, to a great extent, on history.

But assessing communication is not impossible. At a very basic level, speaking skills and listening skills can be assessed by many traditional techniques found in the communication classroom. We have, for many years, had well-established ways of assessing public presentations, interviews, and group meetings. But there are also newer, more experimental technologies that might be harnessed for use in an assessment. Consider a few potential avenues for assessment technology:

> Assessment Centers. One way of gaining an accurate indication of an individual's communication abilities, given its complexity, might be to carefully examine work on assessment centers (Gaugler, Rosenthal, Thorton & Bentson 1987; Joiner 1986). In assessment centers individuals are presented a variety of tasks, often communication related, in groups. Raters evaluate participants on a variety of interpersonal skills. Some typical sorts of group assignments are presented in Appendix B. These are drawn from the AT&T Assessment Center research program summarized by Howard and Bray (1988). The method has a long history of good reliability and validity.
>
> Round-Robin Analyses. In the last few years, methodological techniques for assessing both the independent and interdependent effects of individual interactants in social exchanges have been developed. Statistically, the methods are clear (Warner, Kenny & Soto 1979). Practically, for an assessment, a round-robin procedure would have a group of people have a series of dyadic exchanges among one another. Person A chats with Person B while Person C is chatting with Person D. After some period of time, Person A moves to talk with Person C, Person B with Person D, and so on. Each of these interactions is coded allowing one to separate out what the individual contributes to the conversation.
>
> Videotechnology. New video technologies in the form of such things as interactive videodisks would allow a semblance of interactivity while maintaining some degree of control. Computer-assisted videodisks might be used to adjust to an individual's level of competency in such things as questioning, conflict management, and even the preparation of presentations.

All of these are just examples of ways assessments might be conducted to allow the measurement of important communicator characteristics within the parameters of context and interactivity. More importantly, all offer simultaneous assessments of any number of important communication skills (e.g., listening and speaking can both be assessed at the same time).

Any assessment will need to tap into both typical performance competency and maximal performance capability. What people do everyday and what they are capable of doing are very different things when it comes to oral communication and listening. Everyday communication is often a "mindless" activity (Bavelas & Coates 1992; Kellermann 1992) unless people are seeking a specific goal or the interaction is one that has some degree of unpredictability. Research on listening shows that people's memory for social interactions is amazingly limited (Stafford & Daly 1984; Stafford, Waldron & Infield 1989) unless people are given some incentive to remember. Given an incentive, recall is substantially enhanced. The theme here is straightforward--any assessment needs to examine both typical and maximal performance. The concern for tapping maximal performance also has an educational implication. We want to know not just what people can do in an average setting; we want to know the best they can do. Instruction is best set at a person's maximal level to help them go beyond what they can do.

One way of perhaps measuring maximal performance in communication encounters would be to provide participants with any assessment scenarios that present very difficult communication situations. Difficult encounters, with the appropriate incentives, may challenge people to perform maximally. One way of doing this might be to create "games" where people's communication has consequences for the "points" the receive. Another would be to offer incentives, symbolic or real, to people for accomplishing certain goals. "The more people who are persuaded by your speech, the more X you get." A third might consider communication to be a problem-solving activity. The effective communicator is a better problem solver than the ineffective communicator. One way to tap this problem solving may be by creating tests that challenge respondents to demonstrate their cognitive skill at coping with difficult communication situations. For instance, Daly, Diesel, and Weber (in press) have recently designed a questionnaire that may tap an individual's maximal cognitive skill at problem solving in conversations. They present to subjects "conversational dilemmas." These dilemmas are situations where, at least on the face of it, no matter what a person does, he or she has problems. Beck (1988) provides an example:

Tom: Why are you moping around?

Sally: You told me I was stupid.

Tom: I really didn't mean it. I was angry at the time.

Sally: I know you really do think I'm stupid.

Tom: That just isn't true. I was angry.

Sally: You always say that when people are angry they express their true thoughts.

At this point, Tom faces a conversational dilemma for no matter what he says, he will have a problem. Denying he was angry contradicts what he just said. Suggesting that he was wrong about what he said again contradicts what he has previously said. Saying that Sally misunderstood what he meant is attacking her own sense of understanding. In short, Tom is trapped. Daly, Diesel, and Weber are attempting to identify a number of dilemmas, obtain alternative responses varying in "conversational savvy" as judged by experts, and then rate people on the basis of their choices of the different responses. Their research, to date, suggests that performance on the measure correlates with some important individual differences related to communication performance such as communication apprehension.

Beyond simply focusing on maximal performance it is critical that in any assessment there be incentives for excellent performance. In most "real life" communication settings there are natural incentives: getting the job, impressing the date, keeping others informed, making sure people are happy, and so on. Any valid assessment must include some incentives for performance. Also, it is critical to note the major focus of Goal 5: college graduates. Any assessment must be designed with this population in mind, not high school students, not young adults in the working world, but rather people leaving college after either 2 or 4 years.

Decisions need to be made about the appropriateness of holistic versus more atomistic judgements of communication performance. At the behavior level one might want to judge very specific enactments ("How many times does the interactant interrupt in the conversation? How frequently does this happen") or might prefer to make judgements at a higher level ("Does the interactant allow his or her partner opportunities to interact?"). In public speaking, Goulden (1992) provides examples of each approach as well (see Appendix C). This is one of many practical assessment problems that are not the focus of this paper. But this issue, as well as many others, are important to consider in preparing an assessment for any number of practical, methodological, and conceptual reasons. The Speech Communication Association has devised, after extensive work, some criteria for the assessment of oral communication. It is included in Appendix D.

In any assessment the criteria for judging individuals will be critical. A typical scale might gauge communicators along the following: (1) not knowing what to do, (2) knowing what to do but not how to do it, (3) knowing what to do and how to do it, (4) knowing what to do and how to do it and with whom, (5) knowing what to do, how to do it, with whom, and in what situation. This sort of scaling would be appropriate for some sorts of communication tasks. For other tasks, different methods may be more reasonable. The particular sort of scaling for an assessment is not the focus of this paper. But understand: The scaling will determine the validity of any results.

Decisions need to be made about the actual data point: Will the assessment information focus on an individual's score or will it emphasize group level data? This is a critical concern for what is collected, how it is collected, and how interpretations are to made of the data. Individual-level scores might allow placement of students, might offer individual diagnosis that would aid a participant, and would certainly allow for better empirical studies of variables that might relate to performance. On the other hand, individual-level analysis brings with it a variety of significant problems both in terms of design and sampling. Group-level data will require attention to other issues and will, obviously, limit the potential of the assessment for in-depth analyses, diagnoses, and predictions. The concern goes further: Are we interested in remediation based upon scores? Are we interested in working with students who performs poorly on a test of this sort? If so, the assessment is a very different creature than other national ones. Before much more work is initiated, it will be critical to identify the extent of the assessment.

Assessing Communication Skills: What Should We Assess?

What skills, then, are critical to the college graduate? A number of different alternate schemes are possible. After reviewing many, four seemed particularly reasonable.

One scheme lists major skills in terms of the forms of communication involved (e.g., public speaking, conversing, meeting skills); another lists the various functions communication may serve (e.g., persuasion and compliance gaining, information seeking, giving, and receiving, affinity seeking and maintenance, self-understanding, conflict management (Daly & Wiemann 1993)); another highlights some of the underlying criteria that cut across types of communication and functions (e.g., appropriateness, effectiveness, efficiency, clarity, flexibility and adaptability, involvement and responsiveness). Along with these three one could organize skills along the lines of three very basic dimensions of speaker competence: (1) motivations, beliefs, and predispositions (e.g., shyness, reticence, and communication apprehension (Daly & McCroskey 1984), argumentativeness (Infante & Rancer 1982; Rancer, Kosberg, & Baukus 1992)); (2) knowledge and cognitive indicators (e.g., complexity); and (3) skill enactments or performance (e.g., actual speaking and listening). Taken together, the four schemes yield a four-dimensional structure that may guide any assessment.

These four dimensions provide a potential model for an assessment of communication focusing on work-related and citizenship skills (obviously, in future work it will be imperative to deal with various work-related and citizenship outcomes as another dimension). For exposition purposes we focus primarily on the final three-category system: predispositions and beliefs, knowledge, and performance.

Assessing Motivations and Predispositions to Communicate

Any successful assessment needs to tap individuals' predispositions towards communication. There are some well-established constructs tied to communication-related predispositions for which there are psychometrically acceptable measures. Perhaps the construct most researched within the realm of communication predispositions is a person's shyness, reticence, or communication apprehension. In the past 50 years there have been more than 2,000 studies of this disposition (Daly & McCroskey 1984) and there are more than 40 measures of it in the available literature. Almost all of these measures are self-reports and short forms are available. In virtually every case, the more common measures have highly reliable and acceptable validity. Given the necessity in assessments for brief measures, a short, five-item measure could easily be created to tap general apprehension about communication.

There are, however, other predispositions for which we have fewer or, in a few cases, no available measures. These might include a measure of the perceived utility of communication--a person's belief that communication is a useful activity that can allow him or her to accomplish his or her goals--as well as various socially-oriented personality constructs that are critical predispositions for effective communication performances. These might include argumentativeness (Infante & Rancer 1982), self-monitoring (Synder 1974), assertiveness (Lorr & More 1980), empathy (Davis 1983), adaptability (Duran 1992), conversational sensitivity (Daly, Vangelisti & Daughton 1987), and interaction involvement (Cegala, Savage, Brunner & Conrad 1982).

Assessing Knowledge

This is perhaps the most difficult collection of measures to devise. It is difficult because, in the past, very little research focused on what people should "know" about communication. It might be best to organize this section in an assessment in terms of (1) questions about different strategic accomplishments for which we have some knowledge (e.g., in persuasion, questions about the quality of evidence, the nature of warrants and claims, etc.; similarly, in public speaking there are some "knowns" about visual aids [e.g., a good visual aid has only a limited amount of information on it] as there are in interviewing research [e.g., open questions generally garner longer answers than closed questions]) as well as (2) questions tied to situational descriptors that vary some of the major underlying dimensions of communication encounters (e.g., present a scenario to a respondent where one or the other interactant engages in what most would consider an inappropriate social move. Score the respondent on his recognition of the move as inappropriate). The notion here is that people vary in their abilities to analyze an audience, as do the communication requirements of a setting, as well as audience variables relevant to effectively communicating a message. There is, for instance, a first decision to talk or not to talk; there are role requirements; there are questions any smart speaker might ask about an audience; there is an ability to anticipate what audiences will understand, ask questions about, and react positively or negatively towards in a meeting, conversation, or presentation. In addition, brief, validated, case studies might be used to examine how people organize their understanding of communication events. This last notion provides one means of tapping an important knowledge component of communication--a person's metaknowledge of their communication practices. Just as it is important to know what to do, it is important to know why you are doing what you are doing in a social interaction, meeting, or presentation. Any measure tapping knowledge will need to carefully construct measures of both knowledge and meta-knowledge.

In addition to new techniques such as those mentioned above there are certain "cognitive" measures available in the communication literature that might be appropriate in an assessment. There is considerable work in the field on cognitive complexity (e.g., Burleson 1986; Burleson and Waltman 1988) some of which might be appropriated for an assessment. The evidence is clear that people who are more complex about others as well as about communication encounters (Daly, Bell, Glenn & Lawrence 1985) are more effective communicators than those who are less complex.

Assessing Performance

There are many different forms of communicative performance that may merit assessment in any national project. One issue that will need careful consideration is exactly which forms should be assessed. It will obviously be impractical to assess every form. Perhaps one criteria should be the relevancy of each form to workplace success and citizenship (which forms are most critical to success). Another criteria might be its generalizability (will knowledge of how an individual does on this form generalize to how the individual will perform on other forms). A third might be the practicality of an assessment (any form that appears mostly at home might create massive logistic, ethical, and privacy issues). In this review, for space and time limitations, I have selected just a few forms that meet three criteria: (1) they are especially relevant to workplace settings, (2) there is at least some research about them, and (3) at most colleges and universities, students may receive instruction on them. These forms are public speaking, interviewing, listening, interpersonal or dyadic interactions, and small group decision making. Other forms such as negotiating, conflict resolution, relational communication (e.g., family communication), leadership, delegating, inspiring, coming to understand oneself, computer and telephone exchanges, and visual literacy (e.g., interpreting, understanding, and critiquing media presentations such as television news and advertising) are excluded, not because of their applicability but because of time and space limitations.

Public Speaking Assessment

In assessing public speaking there are a number of skills that are important for effective communicators to demonstrate. The field of speech communication, for the past 60 years, has devised numerous rating forms for classroom use at the college level. Typical ones, drawn from Bock and Bock (1981), are included in Appendix E. From listings such as those, and from other assessments of speaking skills (e.g., Rubin 1982; 1985), the following might be suitable for assessment:

A. The individual's delivery is clear and understandable

   1. Pronunciation and articulation are suitable for the audience and situation

   2. Tone of voice (pitch, intensity, and rate) is varied with appropriate emphases

   3. Nonverbal characteristics (e.g., gestures, facial expressions) are appropriate for the message

   4. Appears comfortable (not nervous)

B. The language the individual uses is effective and appropriate to the situation

C. The speaker demonstrates sufficient knowledge about the topic

D. Sufficient supporting material is provided

E. The presentation is organized in a fashion that allows listeners to understand what is being said, where the speaker is going, with appropriate transitions

1. The introduction is attention-getting, provides a general thesis and an overall summary of the presentation

2. The body of the presentation is well developed, presents issues in an interesting and memorable way, and deals, when appropriate, with counterarguments

3. The conclusion reiterates the key points, provides directions for the audience (when appropriate), and leaves audience members positively disposed towards speaker and topic.

F. The speaker's purpose is clear (although this has some limitations)

G. Handling questions and answers

1. Actively attends to questions

2. Based on listening, correctly understands a question

3. Defends a point of view or position in the face of questions

4. Copes effectively with a "difficult" questioner

H. The speaker accurately adapts

1. To audience characteristics (e.g., size, heterogeneity, interests, concerns)

2. To environmental characteristics (e.g., size of room, equipment, lighting, time)

3. To tools (e.g., notecards, teleprompter, flipcharts)

I. Effective use of visual aids

1. Effectively prepares visual aids that meet typical criteria for a good aid (e.g., not too much information, designed for understandability)

2. Understands which sorts of aids are more or less appropriate in different situations

3. Effectively includes visual aids in presentations

J. Capable of delivering impromptu and extemporaneous speeches as well as preparing and delivering a planned presentation.

Can the ability to give a public presentation be reliably and accurately assessed? The answer is "yes." In classrooms, this has been done for years, sometimes in reliable and valid ways. More systematic attempts have also shown that assessment can be done. Rubin (1982, 1985; Rubin, Graham, & Mignerey 1990), for instance, has demonstrated that with her instrument (any assessment effort needs to carefully look at this research program as an exemplary one) one can obtain valid and reliable indications of people's speaking competency. However, as we've noted elsewhere, the process is expensive, time consuming, and complex. In assessing public speaking it may be reasonable not to assess every skill. There may be a few demonstrable skills that are most critical. Consider a study by Johnson and Szczupakiewicz (1987) who, in large surveys, found that college faculty members felt that informative speaking, persuasive speaking, and gathering supportive material were the top three skills a competent college graduate should demonstrate in speeches. On the other hand, alumni in the workplace felt that informative speaking, listening, and handling questions and answers were key. Both groups agreed that the two most important forms of speeches were extemporaneous and impromptu speeches. One question is often raised when people argue for the assessment of presentation skills: Do people really do it enough for it to be considered important? The response is straightforward: while people may not do it often, when they do it, the consequences are often quite significant.

What might an assessment of public speaking skills look like? Students might be asked to create and then deliver a speech, on some prescaled topic, to an audience (this, by the way, is critical in any assessment--there must be others present and involved). The speech would be rated either "live" or on videotape by trained raters. The raters might judge each of the categories described above ("trait" scales) as well as provide a holistic evaluation of task accomplishment (composed, I would propose, of perceived effectiveness and perceived appropriateness).

Assessing Interviewing Skills

The ability to successfully participate in a job interview is an important skill in the workplace. There are a number of excellent texts on interviewing practices (e.g., Stewart & Cash 1988) and some quite useful research on the topic (e.g., Eder & Ferris 1989; Jablin & Miller 1990). In terms of rating individuals on their performance in interviews a useful measure was devised by Seibold and Meyers (1985). It is reproduced in Appendix F. Something akin to this might be appropriate for assessment.

Listening Assessment

Listening is the flip side of the other performance concerns critical to any assessment of communication skills. It is a receptive skill as opposed to the more productive skills of speaking, interviewing, and meeting with others. Yet it is critical for an accurate assessment of an individual's communication skills if for no other reason than that all the productive skills depend upon excellent receptive skills associated with listening. Various tests of listening have been developed over the last 50 years. It would be impossible, in the space permitted, to summarize all of the research on these tests (see, for reviews, Bostrom 1990; Rubin & Roberts 1987). Some listening research focuses on the display of listening activity (e.g., perceived listening and responsiveness) since that, by itself, is important in social interaction. Other research emphasizes the different sorts of listening tied to recall and interpretation (e.g., short-term listening, long-term listening, interpretative listening) as well as accurate understanding of emotional meaning. Moreover, there has long been serious scholarly interest in the topic of empathic listening (Larson, Backlund, Redmond, & Barbour 1978). In examining the different measures typically used to assess listening (Brown-Carlson Listening Comprehension Test 1955, the Watson-Barker Listening Test 1984, the CCAI Instrument, Rubin & Roberts 1987, Lieb-Brilhart's 1965 measure, and the Kentucky Comprehensive Listening Test, Bostrom 1990) a number of important components for any assessment are notable. They include:

A. The ability to comprehend a message communicated orally in reasonable English

  1. The ability to follow directions and instructions accurately

  2. The ability to accurately understand what was said (meant) by a speaker, including recognition of main ideas, understanding the relationship among ideas, and grasping supporting details presented by the speaker

  3. The ability to correctly synthesize what was said and interpret the implications of the speaker's message

  4. The ability to remember and recall, accurately, what was said after a reasonable time delay

B. The ability to complexly evaluate what was said

  1. Effectively separating fact from emotion

  2. Effectively separating evidenced material from unevidenced material

  3. Effectively detecting bias and prejudice

C. The ability to evaluate the emotional meaning of the message and speaker

    1. The ability to accurately detect the emotions and feelings of the speaker

    2. The ability to accurately infer power and affinity relationships among interactants (predominantly in meeting interactions)

D. The ability to nonverbally display indications of listening and involvement

    1. To indicate perceived listening

    2. To understand when it is most appropriate to "say nothing" and simply attend

    3. To demonstrate successful turn-taking in conversation (i.e., avoid interrupting but know when it is time to talk)

E. The ability to successfully make use of material gained through listening in further communication encounters

    1. In conversations, to engage in coherent discourse based upon what was said previously in the exchange

    2. In meetings, to incorporate others' statements into one's own statements

    3. In presentations to accurately understand questions raised by audience members.

The assessment of listening, on the face of it, seems much easier than the assessment of speaking. This, upon analysis, may not be the case. Consider the following issues as but examples:

First, in any assessment, generalization depends upon stimuli presentation. If audio tapes are used, the generalization will be to a very limited sample of situations; if video is used, again, the generalization to typical interactions may be questionable. Certainly both of these stimuli presentations represent some sorts of communication events (e.g., radio, television, telephone), but much interaction happens face-to-face. Thus, at least part of any assessment must include face-to-face exchanges where listening is assessed. Closely related is the environment in which any assessment is completed. For instance, typical communication events are filled with distractions. Assessments often attempt to control distractions. Yet doing this may not offer a valid test of how well an individual listens in normal settings. Yet creating distractions may cause as many problems as not having them.

Second, to the extent that face-to-face exchanges are used, control over issue involvement, topic knowledge, and other related concerns becomes an issue. In a face-to-face conversation and meetings participants are listeners and speakers. What, then, is the critical recall measure: what the other person(s) said, what the person being assessed said, or what was said in the conversation regardless of who uttered it?

Third, in assessing recall, what to assess becomes a critical concern. There have been a number of coding schemes devised to assess recall of social interaction, starting with work by Stafford and Daly (1984). Most research has focused on recall of something akin to "idea units," eschewing direct recall of specific words and sentences. Deciding what the critical dependent variable is in a listening assessment may well depend upon what one wishes to generalize to. In some work settings, specific recall is critical. In some conversational situations, specific recall is far less important. And, as research has demonstrated, sometimes the most interesting dependent variable is not what the person accurately recalls but what the person recalls that wasn't said in the interaction, speech, or meeting.

Fourth, while everyone would agree that accurate interpretation of the speaker's meaning and emotions is a critical characteristic of good listening, being able to posit what a speaker meant or felt ahead of time, especially in conversational and meeting interactions, will be difficult.

Assessing Conversational Performance

What needs to be assessed in tapping an individual's performance competencies in social interactions? Hints exist throughout communication literature ranging from Wiemann's (1977) broad foray where he identified three major components: empathy, behavioral flexibility, and interaction management to Spitzberg's (1989) much more extensive discussion. One useful coding scheme is one proposed by Spitzberg and Hurt (1987). It is reproduced in Appendix G.

If one were to categorize effective interpersonal skills the following might be included.

A. Appropriate empathy for others involved in the exchange

B. Appropriate displays of involvement and supportiveness

    1. Showing attention and responsiveness

    2. Memory for the interaction

    3. Using what the other said as a means of continuing the exchange

C. Conversational and interaction management skills

    1. Appropriate turn taking

    2. Keeping the conversation going (no "deadtimes")

    3. Appropriate opening, leave taking (Kellermann, Reynolds & Chen 1991)

  4. Maintain coherence (Craig & Tracy 1983)

D. Express thoughts and emotions clearly and appropriately

E. Encourage and generate interaction from other

  1. Seek information

  2. Get other to disclose

F. Understand what other is saying

  1. Grasp implicature (Grice 1975)

  2. Aid other in clarifying thoughts

G. Adaptability (e.g., Duran 1992)

  1. Adapts effectively to situation

  2. Adapts effectively to topic(s)

  3. Adapts effectively to other interactants

H. Cope with difficult communication events

  1. Effectively resist, in socially appropriate ways, attempt to persuade (e.g., Lim 1990)

  2. Successfully detect deception (e.g., McCornack 1992; deTurck et al. 1990)

  3. Respond in socially acceptable ways to troublesome communicators

  4. Successfully handle interpersonal conflict situations

  5. Effectively persuade

  6. Identify and resolve misunderstandings

Interpersonal effectiveness, in varied forms, has long been of interest to scholars in communication. People have studied social interaction within settings involving persuasion and compliance gaining (e.g., Cody & McLaughlin 1990; Dillard 1990), conflict resolution (Donohue & Kolt 1992; Hocker & Wilmot 1991), conversational processes (e.g., Nofsinger 1991), relational communication (Knapp & Vangelisti 1992), and communication in marriage

(Fitzpatrick & Noller 1988) among many others. The list of characteristics is but a sampling drawn from those literatures.

When completing an assessment of interpersonal effectiveness the researcher will face a number of problems, conceptual and methodological in nature. For instance, there is a level of analysis problem. Is the appropriate level the effective communicator or the effective conversation? There is the issue of acquaintanceship: Is the assessment made with people who are well acquainted (this is a must for some sorts of behaviors and for generalization to some sorts of tasks and settings) or with people who are unacquainted (initial interaction skills are critical to any future exchanges). There are also numerous other concerns about issues such as context, goals, incentives, coding of data, and inferences made from talk to mental models. While assessments are possible, the difficulties inherent in them will challenge even the best researchers.

Assessing Small Group Exchanges

In work settings people spend more and more of their time working in teams and groups to accomplish their goals. What skills are critical for effective team membership? A review of the literature suggests that the skills fall into three categories: (1) individual level skills, (2) group characteristics (e.g., appropriate heterogeneity, group-generated goals), and (3) environmental or context variables (e.g., support from supervisors, appropriate materials and tools, assigned purpose). This isn't the place to review all the available literature (e.g., Brown 1990; Dyer 1987; Larson & LaFasto 1989) about these categories but any assessment is going to have to grapple with the three different categories. For purposes of discussion, let's limit the criteria to individual-level skills. They might include

A. The ability to stay involved in the discussion

B. A consistent focus on the purpose or goal of the exchange

C. Communicating (participating) effectively

    1. Initiates interaction
    2. Uses summary statements (Leathers 1972)
    3. Establishes relationships among ideas
    4. Avoids high-level inferences
    5. Makes relevant contributions: Avoids irrelevancies
    6. Tests assumptions

   7. Identifies problems and issues
   8. Generates alternatives
   9. Stimulates thinking
   10. Listens to others

D. Maintaining leadership
   1. Focuses on processes that enable the group (Hirokawa (1983)
   2. Makes topic changes in a supportive manner
   3. Seeks clarification and elaboration
   4. Manages group time

E. Being well prepared

F. Coping with problems
   1. Manages troublesome members
   2. Keeps group focused
   3. Controls conflict and hostility
   4. Generates discussion from quiet members
   5. Includes others

G. Decision making (Gouran 1988)
   1. Correctness
   2. Quality
   3. Utility
   4. Acceptability

There are clearly other components of effective group participation. The list is preliminary in nature. However, before much work is completed on the assessment of group interaction, it is critical to determine whether the best focus is on the individual, the group, or both. Also important is a consideration of the "reality" of the interaction. While there are any number of exercises one might use to assess individual and group performance, it is essential that any stimulus have external validity to workplace teams.

Assessment Criteria

Perhaps the most parsimonious way of organizing communication competencies that merit assessment might be to tap into the underlying competencies one would anticipate in every interaction, whether it be a conversation, a meeting, an interview, or a public presentation. Every communicator must demonstrate:

A. Situational appropriateness: In every interaction, every speech, every meeting there are certain behaviors that are less appropriate than others. A skilled communicator can "read" the situation, know what is appropriate and what is inappropriate, and behave accordingly.

B. Involvement and responsiveness: Effective communication requires, under most circumstances, that every participant be involved and responsive. Involvement and responsiveness are partly nonverbal activities and partly verbal.

C. Adaptability and flexibility: Good communicators adapt to other interactants, demands of the setting, and changing contingencies in the interaction. They are flexible in approaching interaction problems.

D. Clarity: Under most circumstances (when ambiguity and equivocation [see Bavelas, Black, Chovil & Mullett 1990] are not strategic), good communicators are clear in communicating their messages. Not only are messages understandable in terms of language and delivery, but they are clear in terms of intent.

E. Efficiency: Good communicators are efficient in communicating their messages (Kellerman et al. 1991). They don't waste the time and efforts of others or themselves on unnecessary moves, useless exercises, and tedious verbiage.

F. Goal Accomplishment: Effective communicators, in purposeful interactions, know what they want to accomplish and go about attempting to accomplish that goal. Obviously, goal accomplishment interacts with all the components of this section. Selecting an appropriate goal within a context and performing the appropriate behaviors appropriately are key notions. Indeed, one respondent to this paper (Friedrich) notes that a potentially useful definition of communication would be one couched within goal selection: "A rough preliminary definition might define communication competence as a situational ability to set realistic and appropriate goals and to maximize their achievement by using knowledge of self, other, context, and communication theory to generate adaptive communication performance."

G. Politeness: There are face-saving rules (Brown & Levinson 1987) in any interaction that must be maintained. An effective interactant understands these rules and engages in them as appropriate.

When the different criteria are closely examined, one could easily argue that there are two major continua most communication events might be judged upon in a rating situation: effectiveness and appropriateness. All six of the criteria described above fit into these two very basic dimensions. Under the rubric of effectiveness would be variables such as clarity, memorability, impact, coherence, and efficiency. Under the label appropriateness are terms such as adaptability, flexibility, relevance, and awareness of social norms.

Functions of Communication

For the past 15 years communication scholars have carefully identified different sorts of goals communicators might have when communicating with others, regardless of whether the communication event is a speech, a conversation, a meeting, or an interview. Following broadly under the rubric of strategies (Daly & Wiemann 1993) this research has carefully considered the different moves people might make when attempting to accomplish their goals. While the list of goals would obviously be potentially endless, a few major ones emerge in the literature. They include (1) influence and compliance gaining as well as resisting influence attempts, (2) information seeking and giving, (3) affinity seeking and maintenance, (4) conflict management, (5) expressing (e.g., self-disclosure) (6) explanations and accounts, (7) coping strategies for problematic communication events (e.g., defensiveness, embarrassment), and (8) "negative" behaviors that, nonetheless, mark effectiveness (e.g., inspiring guilt, using anger effectively). One assessment strategy that might cut across forms of communication would examine the effectiveness of individuals at achieving these various goals through communication while at the same time meeting the criteria described previously (e.g., politeness, appropriateness, flexibility, effectiveness).

When scholars spend time discussing communication in terms of its function, the discussion usually concludes that there are three major communicative functions: to inform, to persuade, and to develop and maintain relationships. These three functions judged along the two continuua of appropriateness and effectiveness provide a useful schematic representation for communication evaluation, regardless of the particular skill being assessed or the specific context within which the assessment is taking place. Actually, it is probably better to consider all three functions as relevant, to varying degrees, in any communication encounter as are the two criteria of effectiveness and appropriateness. In any conversation, meeting, interview, or presentation people want to provide some amount of information that is understood and remembered, want, to some extent to persuade, and want, in some ways to develop or maintain a relationship. The quality of the performance is judged in terms of how effectively the person accomplishes these goals and how appropriate his or her enactments are to the context, topic, and people involved in the exchange.

References

Alexander, E., Penley, L. & Jernigan, I.E. (1992). The relationship of basic decoding skills to managerial effectiveness. *Management Communication Quarterly*, 6, 58-73.

Bavelas, J., et al. (1990). *Equivocal communication*. Newbury Park: SAGE.

Bavelas, J.B. & Coates, L. (1992). How do we account for mindfulness of face-to-face dialogue? *Communication Monographs*, 59, 301-305.

Beck, A. (1988). *Love is never enough*. NY: Harper & Row.

Bock, D.G. & Bock, E.H. (1981). *Evaluating classroom speaking*. Urbana: ERIC.

Bostrom, R., (1990). *Listening behavior: Measurement and application*. New York: Guilford.

Brown, R. (1990). *Group processes: Dynamics within and between groups*. London: Blackwell.

Brown, P. & Levinson, S.C. (1987). *Politeness: Some universals in language*. NY: Cambridge.

Brown, J. & Carlson, G. (1955). *Brown-Carlson Listening Comprehension Test, Manual of Directions*. Yonkers: World Book.

Burleson, B. (1986). Cognitive complexity. In J.C. McCroskey & J.A. Daly (Eds.), *Personality and interpersonal communication*. Newbury Park: SAGE.

Burleson, B. & Waltman, M.S. (1988). Cognitive complexity: Using the role category questionnaire. In C.H. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes*. New York: Ablex.

Cegala, D.J., Savage, G.T., Brunner, C.C. & Conrad, A.B. (1982). An elaboration of interaction involvement: Toward the development of a theoretical concept. *Communication Monographs*, 49, 229-248.

Cody, M. & McLaughlin, M. (1990). *The psychology of tactical communication*. Avon, UK: Multilingual Matters.

Craig, R.T. & Tracy (Eds.). (1983). *Conversational coherence: Form, structure, and strategy*. Beverly Hills: SAGE.

Curtis, D.B., Winsor, J.L. & Stephens, R.D. (1989). National preferences in business and communication education. *Communication Education*, 38, 6-14

Daly, J.A., Bell, R., Glenn, P. & Lawrence, S. (1985). Conceptualizing communication complexity. *Human Communication Research*, 12, 30-53.

Daly, J.A., Diesel, C. & Weber, D. (in press). Conversational dilemmas. In W. Cupach & B. Spitzberg (Eds.), *The dark side of interpersonal communication*. Hillsdale, NJ: Erlbaum.

Daly, J.A. & McCroskey, J.C. (1984). *Avoiding communication: Shyness, reticence, and communication apprehension*. Beverly Hills: SAGE.

Daly, J.A., Vangelisti, A.L. & Daughton, S. (1987). The nature and correlates of conversational sensitivity. *Human Communication Research*, 14, 167-202.

Daly, J.A. & Wiemann, J. (1993). *Communicating strategically*. Hillsdale, NJ: Erlbaum.

Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126.

deTurck, M.A., Harszlak, J., Bodham, D., & Texter, L. (1990). The effects of training social perceivers to detect deception from behavioral cues. *Communication Quarterly*, 38, 189-199.

Dillard, J. (1990). *Seeking compliance: The production of interpersonal influence messages*. Scottsdale, AZ: Gorsuch Scarisbrick.

DiSalvo, V.S. (1980). A summary of current research identifying communication skills in various organizational contexts. *Communication Education*, 29, 283-290.

DiSalvo, V.S., Larsen, D. & Seiler, W. (1976). Communication skills needed by persons in business organizations. *Communication Education*, 25, 273-275.

Donohue, W. & Kolt, R. (1992). *Managing interpersonal conflict*. Newbury, CA: SAGE.

Duck, S. & Silver, R. (1990). *Personal relationships and social support*. London: Sage.

Duran, R.L. (1992). Communication adaptability: A review of conceptualization and measurement. *Communication Quarterly*, 40, 253-268.

Dyer, W. (1987). *Teambuilding: Issues and alternatives*. Reading, MA: Addison-Wesley.

Eder, R.W. & Ferris, G.R. (1989). *The employment interview: Theory, research, and practice*. Newbury Park, California: Sage.

Fitzpatrick, M. & Noller, P. (1988). *Between husbands and wives: Communication in marriage*. Newbury Park: SAGE.

Gaugler, B.B., Rosenthal, D.B., Thorton, G,C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493-511.

Gottman, J. (1979). *Marital interaction: Experimental investigations*. NY: Academic Press.

Goulden, N.R. (1992). Theory and vocabulary for communication assessment. *Communication Education*, 41, 258-269.

Gouran, D. (1983). Communicative influences on inferential judgements in decision-making groups: A descriptive analysis. In D. Zarefsky et al. (Eds.) *Argument in Transition*. Annandale, VA: Speech Communication Association 667-684.

Gouran, D. (1988). Group decision making: An approach to integrative research. In C.H. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes*. New York: Ablex.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*, (Vol. 3). New York: Academic Press.

Hanna, M.S. (1978). Speech communication training needs in the business community. *Central States Speech Journal*, 29, 163-172.

Hay, E.A. (1992). A national survey of assessment trends in communication departments. *Communication Education*, 41, 247-257.

Henry, J.J. & Richmond, S.U. (1982). *Basic skills in the United States work force*. New York: Center for Public Resources.

Hirokawa, R.Y. (1983). Group communication and problem solving effectiveness: An investigation of group phases. *Human Communication Research*, 9, 291-305.

Hirokawa, R.Y. (1988). Group communication research: Considerations for use of interaction analysis. In C.H. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes*. NY: Ablex.

Hirokawa, R. & Pace, R. (1983). A descriptive investigation of the possible communication-based reasons for effective and ineffective group decision making. *Communication Monographs*, 50, 363-379.

Hocker, J. & Wilmot, W. (1991). *Interpersonal conflict*. Dubuque, IA: Wm. C. Brown.

Howard, A. & Bray, D.W. (1988). *Managerial lives in transition: Advancing age and changing times*. New York: Guilford.

Huston, T. & Vangelisti, A. (1991). Socio-emotional behavior and satisfaction in marital relationships: A longitudinal study. *Journal of Personality and Social Psychology*, 61, 721-733.

Infante, D. & Rancer, A. (1982). A Conceptualization and measurement of argumentativeness. *Journal of Personality Assessment*, 46, 72-80.

Jablin, F.M. & Miller, V.D. (1990). Interviewer and applicant questioning behavior in employment interviews. *Management Communication Quarterly*, 4, 51-86.

Joiner, D.A. (1986). Assessment centers in the public sector: A practical approach. *Public Personnel Management Journal*, 13, 435-450.

Johnson, J.R. & Szczupakiewicz, N. (1987). The public speaking course: Is it preparing students with work related public speaking skills? *Communication Education*, 36, 131-137.

Kellermann, K. (1992). Communication: Inherently strategic and primarily automatic. *Communication Monographs*, 59, 288-300.

Kellermann, K., Reynolds, R. & Chen, J. (1991). Strategies of conversational retreat: When parting is not sweet sorrow. *Communication Monographs*, 58, 362-383.

Klemmer, E.T. & Synder, F.W. (1972). Measurement of time spent communicating. *Journal of Communication*, 20, 142-158.

Knapp, M. L. & Hall, J. (1992). *Nonverbal communication in human interaction*. New York: Holt, Rinehart & Winston.

Knapp, M.L. & Vangelisti, A.L. (1992). *Interpersonal communication and human relations*. Boston: Allyn & Bacon.

Larson, C., Backlund, P., Redmond, M. & Barbour, A. (1978). *Assessing functional communication*. Urbana, IL: ERIC.

Larson, C. & LaFasto, F. (1989). *Teamwork: What must go right, What can go wrong*. Newbury Park, CA: SAGE.

Leathers, D. (1972). Quality of group communication as a determinant of group product. *Speech Monographs*, 39, 166-173.

Lieb-Brilhart, B. (1965). The relationship between some aspects of communicative speaking and communicative listening. *Journal of Communication*, 15, 35-46.

Lim, T. (1990). The influence of receivers' resistance on persuaders' verbal aggressiveness.

Communication Quarterly, 38, 170-188.

Lorr, M. & More, W.W. (1980). Four dimensions of assertiveness. *Multivariate Behavioral Research*, 2, 127-138.

Luthens, F., Hodgetts, R. & Rosenkrantz, S. (1988). *Real Managers*. Cambridge: Ballinger.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs*, 59, 1-16.

National Center for Educational Statistics. (1992). *Assessing literacy: The framework for the National Adult Literacy Survey*. Washington DC: U.S. Department of Education.

Nofsinger, R. (1991). *Everyday conversation*. Newbury Park: SAGE.

Noller, P. & Fitzpatrick, M. (1990). Marital communication in the eighties. *Journal of Marriage and the Family*, 52, 832-843.

Ong, W. (1982). *Orality and literacy: The technologizing of the word*. New York: Methuen.

Papa, M.J. & Graham, E.E. (1991). The impact of diagnosing skill deficiencies and assessment-based communication training on managerial performance. *Communication Education*, 40, 368-384.

Rancer, A.S., Kosberg, R.L. & Baukus, R.A. (1992). Beliefs about arguing as predictors of trait argumentativeness: Implications for training in argument and conflict management. *Communication Education*, 41, 375-386.

Rankin, P. (1928). The importance of listening ability. *English Journal*, 17, 623-630.

Rosenthal, R., Hall, J., DiMatteo, M., Rogers, P. & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: Johns Hopkins University Press.

Rubin, R.B. (1982). Assessing speaking and listening competency at the college level: The communication competency assessment instrument. *Communication Education*, 31, 9-18.

Rubin, R.B. (1985). The validity of the communication competency assessment instrument. *Communication Monographs*, 52, 173-185.

Rubin, R.B. & Roberts, C.V. (1987). A comparative examination and analysis of three listening tests. *Communication Education*, 36, 142-153.

Rubin, R.B., Graham, E.E., & Mignerey, J.T. (1990). A longitudinal study of college students' communication competence. *Communication Education*, 39, 1-14.

Seibold, D. & Meyer, R. (1985). Co-participant perceptions of information gathering interviews: Implications for teaching interviewing skills. *Communication Education*, 34, 106-118.

Snyder, M. (1974). The self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 526-537.

Spitzberg, B. (1989). *Handbook of interpersonal competence research*. NY: Springer-Verlag

Spitzberg, B. & Cupach, W. (1984). *Interpersonal communication competence*. Beverly Hills: SAGE.

Spitzberg, B. & Brunner, C. (1991). Toward a theoretical integration of context and competence inference research. *Western Journal of Speech Communication*, 55, 28-46.

Spitzberg, G. & Hurt, H.T. (1987). The measurement of interpersonal skills in instructional contexts. *Communication Education*, 36, 28-45.

Stafford, L. & Daly, J. (1984). Conversational memory: The effects of recall and memory expectancies on remembrances of natural conversations. *Human Communication Research*, 10, 379-402.

Stafford, L., Waldron, V. & Infield, L. (1989). Actor-observer differences in conversational memory. *Human Communication Research*, 15, 590-611.

Stewart, C.J. & Cash, W.B. (1988). *Interviewing: Principles and practices*, (5th Ed.). Dubuque, IA: Wm. Brown.

Sypher, B.D. & Zorn, T.E. (1986). Communication-related abilities and upward mobility. *Human Communication Research*, 12, 420-431.

Vangelisti, A.L. & Daly, J.A. (1989). Correlates of speaking skills in the United States: A national assessment. *Communication Education*, 38, 132-143.

Warner, R,M., Kenny, D.A. & Soto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, 37, 1742-1757.

Watson, K.W. & Barker, L.L. (1984). Listening behavior: definition and measurement. In R.N. Bostrom (Ed.), *Communication Yearbook*, 8, (pp. 178-197). Beverly Hills: Sage.

Wiemann, J.M. (1977). Explication and test of a model of communicative competence. *Human Communication Research*, 3, 195-213.

Wolvin, A. & Coakley, C. (1991). A survey of the status of listening training in some Fortune 500 corporations. *Communication Education*, 40, 152-164.

Appendix A

Summary Tables (from Curtis, Winsor & Stephens 1989)

Factors Most Important in Helping Graduating College Students Obtain Employment

Rank/Order	Factors/Skills Evaluated	Score
1	Oral (speaking) communication	4.566
2	Listening ability	4.232
3	Enthusiasm	4.218
4	Written communication sills	4.162
5	Technical competence	4.157
6	Appearance	3.986
7	Poise	3.859
8	Work experience	3.843
9	Resume	3.612
10	Specific degree held	3.412
11	Grade point average	3.371
12	Part-time or summer employment	3.339
13	Accreditation of program	3.258
14	Leadership in campus/community activities	3.152
15	Participation in campus/community activities	3.058
16	Recommendations	2.965
17	School Attended	2.824

Factors/Skills Important for Successful Job Performance

Rank/Order	Factors/Skills Rated Important	Score
1	Interpersonal/human relations skills	4.480
2	Oral (speaking) communications skills	4.462
3	Written communication skills	4.278
4	Persistence/determination	4.198
5	Enthusiasm	4.164
6	Technical competence	4.120
7	Personality	3.853
8	Work experience	3.769
9	Dress/grooming	3.727
10	Poise	3.719
11	Interviewing skills	3.420
12	Specific degree held	2.925
13	Physical attractiveness	2.746
14	Grade point average	2.682
15	Resume	2.528
16	School attended	2.393
17	Letters of recommendation	2.344

IDEAL MANAGEMENT PROFILE

Rank	Trait/Skill	Score
1	Ability to work well with others one-on-one	4.533
2	Ability to gather accurate information from others to make a decision	4.461
3	Ability to work well in small groups	4.452
4	Ability to listen effectively and give counsel	4.437
5	Ability to give effective feedback (appraisal)	4.262
6	Ability to write effective business reports	4.257
7	Knowledge of job	4.153
8	Ability to present a good public image for the organization	4.071
9	Ability to use computers	3.696
10	Knowledge of management theory	3.425
11	Knowledge of finance	3.342
12	Knowledge of marketing	3.311
13	Knowledge of accounting	3.210
14	Ability to use business machines	2.861

Appendix B

Assessment Center Tasks (from Howard & Bray 1988)

Managerial Simulations

Business Game

The business game at the MPS:O assessment center was called the "Manufacturing Problem." This simulation, in which six subjects at a time participated together, involved purchasing Tinker Toy parts, using them to manufacture simple toys (models of which were provided), and selling them back to the member of the assessment staff designated as the buyer/supplier. The problem was divided into six periods for a total of approximately 2 hours. Prices for products and amounts for which assembled toys could be sold changed sharply from period to period, and good organization was necessary to show a profit. Group action was necessary; no one could buy or sell as an individual. The group was given a small amount of money to start with and was allowed to keep any profit made.

The business game used at MPS:8 was the "Investment Problem." This simulation, in which six subjects at a time participated together, involved a miniature stock market (listing 15 fictitious issues). The participants played the roles of managing partners of a mutual fund. Their task was to invest the fund's money (bogus) to make a profit for the shareholders. Stock prices fluctuated during the several trading periods in accordance with general market trends and factors affecting particular issues. Group action was necessary; no one could buy or sell as an individual.

No business game was included in the MCS assessment exercises.

Group Discussion

At MPS:O, the group discussion exercise was the "Promotion Problem." Again in groups of six, the participants were instructed to assume that they were second-level managers meeting to decide which of their subordinates should be promoted to fill a current vacancy. Each participant was told that he had a strong candidate from among his own subordinates and was given a 300-word summary of this person's strengths and weaknesses. Each participant was required to make a short formal presentation of his candidate. The group then was given 1 hour for free discussion to select one of the six foremen to receive the promotion and to rank-order the others.

The group discussion problem used in the MPS:8 centers was the "Organization Problem." In this, groups of six participants were asked to assume they were a task force preparing recommendations for top management on how the telephone company might be reorganized for greater efficiency. As background, they were each given a copy of a recent interviewing study on possible reorganization and allowed 1 hour for preparation. Each participant was then allowed up to 10 minutes to present his recommendations orally. A 1-hour discussion followed in which the participants were required to agree on proposals for top management.

There were two group discussions in MCS assessment. One of these paralleled the original MPS Promotion Problem in that it was structured to produce competition. The "City Council Problem" (made available gratis by Development Dimensions International, the copyright owner) had each group of six participants play the roles of members of the city council of a small Midwestern city. They were told that they were about to attend a meeting of the council in which they would have to decide how to spend a $1,000,000 windfall from the federal government. Each participant was given the same background data on the fictitious city and its budget. Each participant was told that he or she would represent a different city department (such as the Fire or Police Department) and that he or she should attempt to secure at least a major share of the money for that department. The participants were given 30 minutes to prepare and 10 minutes to present their proposals at the council meeting. A 1-hour discussion followed, with the requirement that the group agree on the distribution of $1,000,000 before time elapsed.

The 119 participants assessed during the first 2 MCS years experienced a different but highly comparable exercise, the "School Board Problem." Here the task was also the distribution of a budget windfall, this time to particular aspects of school district functioning. This exercise was discontinued due to its adoption in an operational assessment center program that some of the participants might later attend.

The second MCS group discussion exercise, used throughout MCS assessment, was the "Counseling Center Group Discussion," originally developed for MPS:20. In this, six assessees were asked to assume that they were senior members of a community counseling center meeting to discuss guidance for junior counselors dealing with five different clients, each with a different "problem of living." They were given brief descriptions of each case and required to develop a consensus on guidance to be given the junior counselors.

At the conclusion of each group exercise, overall ratings and rankings of individual performance were made by the observers and by the participants themselves.

Appendix C

Typical Speech Rating Forms Emphasizing Atomistic and Holistic Ratings (from Goulden 1992)

Excerpt from Analytic Instrument for Persuasive Speech

Instructions: Rate and mark each characteristic using the following scale. Add the subscores together for the speech score.

Excellent-4; Good-3; Satisfactory-2; Inadequate-1; Absent-0.

1. Thesis tells what audience should believe and/or do. _____
2. Speaker shows problem is widespread and/or severe. _____
5. Speaker employs natural deliver style. _____
12. Speaker uses language choices that fulfill purpose. _____
20. Main points are organized in a persuasive pattern. _____

Excerpt from Holistic Instrument for Pubic Speaking

Instructions: Select the level which most closely matches the speech. If the description is consistent throughout with the presentation, then points near the top of the range for that level should be assigned.

30-21 Points
Excellent Level: Topic and materials accurately tailored to audience needs, interest, and knowledge level. Speaker shared logical structure of speech so that audience could effortlessly follow organization. The body fully developed the thesis; support materials were relevant, current, sufficient, clear. Speaker included original creative elements which set the speech apart as unique. Delivery was extemporaneous, natural and confident. Voice, body movements, and eye contact established and maintained effective connection with audience. Speaker refrained from distracting behaviors.

10-0 Points
Poor Level: Topic and materials did not fit audience as they were too technical, too trivial, too familiar or did not apply to audience needs or interest. Speaker either did not have a plan of organization or failed to share the plan with audience. Support materials lacked relevancy, currency, clarity, or variety. Mode of delivery made it difficult for speaker to connect with audience because speech was read, or memorized, or speaker lacked fluency. Speaker's body was stiff, unnaturally static or displayed excessive movement. Eye contact was not sustained with audience.

Appendix D

The Speech Communication Association's Criteria for the Assessment of Oral Communication

The Speech Communication Association's Criteria for the Assessment of Oral Communication

General Criteria

1. Assessment of oral communication should view competence in oral communication as a gestalt of several interacting dimensions. At a minimum, all assessments of oral communication should include an assessment of knowledge (understanding communication process; comprehension of the elements, rules, and dynamics of a communication event; awareness of what is appropriate in a communication situation), an assessment of skills (the possession of a repertoire of skills and the actual performance of skills), and an evaluation of the individual's attitude toward communication (e.g., value placed on oral communication, apprehension, reticence, willingness to communicate, readiness to communicate).

2. Because oral communication is an interactive and social process, assessment should consider the judgment of a trained assessor as well as the impressions of others involved in the communication act (audience, interviewer, other group members, conversant), and may include the self-report of the individual being assessed.

3. Assessment of oral communication should clearly distinguish speaking and listening from reading and writing. While some parts of the assessment process may include reading and writing, a major portion of the assessment of oral communication should require speaking and listening. Directions from the assessor and responses by the individual being assessed should be in the oral/aural mode.

4. Assessment of oral communication should be sensitive to the effects of relevant physical and psychological disabilities on the assessment of competence (e.g., with appropriate aids in signal reception, a hearing impaired person can be a competent empathic listener).

5. Assessment of oral communication should be based in part on atomistic/analytic data collected and on a holistic impression.

Criteria for the Content of Assessment

1. Assessment of oral communication for all students should include assessment of both verbal and nonverbal aspects of communication and should consider competence in more than one communication setting. As a minimum, assessment should occur in the one-to-many setting (e.g., public speaking, practical small group discussion) and in the one-to-one setting (e.g., interviews, interpersonal relations).

2. Assessment of speech majors and other oral communication specialists could include in addition assessment in specialized fields appropriate to the course of study followed or the specialty of the person being assessed.

Criteria for Assessment Instruments

1. The method of assessment should be consistent with the dimension of oral communication being assessed. While knowledge and attitude may be assessed in part through paper and pencil instruments, speaking skills must be assessed through actual performance in social settings (speaking before an audience, undergoing an interview, participating in a group discussion, etc.) appropriate to the skills being assessed.

2. Instruments for assessing oral communication should describe degrees of competence. Either/or descriptions such as "competent" or "incompetent" should be avoided as should attempts to diagnose reasons why individuals demonstrate or fail to demonstrate particular degrees of competence.

3. Instruments for assessing each dimension of oral communication competence should clearly identify the range of responses which constitute various degrees of competence. Examples of such responses should be provided as anchors.

4. Assessment instruments should have an acceptable level of reliability (e.g., test/retest reliability, split-half reliability, alternative forms reliability, inter-rater reliability, and internal consistency).

5. Assessment instruments should have appropriate validity: content validity, predictive validity, and concurrent validity.

6. Assessment instruments must meet acceptable standards for freedom from cultural, sexual, ethical, racial, age, and developmental bias.

7. Assessment instruments should be suitable for the developmental level of the individual being assessed.

8. Assessment instruments should be standardized and detailed enough so that individual responses will not be affected by an administrator's skill in administering the procedures.

Criteria for Assessment Procedures and Administration

1. Assessment procedures should protect the rights of those being assessed in the following ways: administration of assessment instruments and assessment and the uses

of assessment results should be kept confidential and be released only to an appropriate institutional office, to the individual assessed, or if a minor, to his or her parent or legal guardian.

2. Use of competence assessment as a basis for procedural decisions concerning an individual should, when feasible, be based on multiple sources of information, including especially (a) direct evidence of actual communication performance in school and/or other contexts, (b) results of formal competence assessment, and (c) measures of individual attitudes toward communication (e.g., value placed on oral communication, apprehension, reticence, willingness to communicate, and readiness to communicate).

3. Individuals administering assessment procedures for oral communication should have received sufficient training by speech communication professionals to make their assessment reliable. Scoring of some standardized assessment instruments in speaking and listening may require specialized training in oral communication on the part of the assessor.

Criteria for Assessment Frequency

Periodic assessment of oral communication competency should occur annually during the educational careers of students. An effective systematic assessment program minimally should occur at educational levels K, 4, 8, 12, 14, and 16.

Criteria for the Use of Assessment Results

The results of student oral communication competency assessment should be used in an ethical, nondiscriminatory manner for such purposes as:

1. Diagnosing student strengths and weaknesses;

2. Planning instructional strategies to address student strengths and weaknesses;

3. Certification of student readiness for entry into and exit from programs and institutions;

4. Evaluating and describing overall student achievement;

5. Screening students for programs designed for special populations;

6. Counseling students for academic and career options; and

7. Evaluating the effectiveness of instructional programs.

No single assessment instrument is likely to support all these purposes. Moreover, instruments appropriate to various or multiple purposes typically vary in length, breadth/depth of content, technical rigor, and format.

*The criteria contained in this document were originally adopted as resolutions at the SCA Conference on Assessment in Denver, Colorado, in July of 1990. Several of the criteria were authored by the Committee on Assessment and Testing Subcommittee on Criteria for Content, Procedures, and Guidelines for Oral Communication Competencies composed of James W. Crocker-Lakness (Subcommittee Chair), Sandra Manheimer, and Tom E. Scott. The introductions entitled "A National Context" and "SCA's Assessment Activities," were authored by James W. Chesebro, SCA Director of Education Services.

For additional information regarding the assessment of communication, contact,

Speech Communication Association
5105 Backlick Road, Building E
Annandale, Virginia 22003
Telephone (703) 750-9533
FAX (703) 914-9471

Appendix E

Typical Speech Rating Forms (from Bock & Bock 1981)

#10 Speech Performance Scale

Name of speaker____Date________
Evaluator____Section____________________
Speech Topic____

Instructions: Please rate each of the six categories below on the following 1 to 5 scale: 1 equals poor, 2 equals fair, 3 equals average, 4 equals good, 5 equals superior. Use the criteria under each of the six categories as a basis for your rating. Also rate the speech on General Effectiveness in the special box at the bottom of this page.

Criteria	Rate 1-5
Delivery: Easily audible — Good pitch — Varied Not forced — Pleasant quality — Conversational Good rate — Fluent — Acceptable Clear — Responsive — Well controlled Adaptive	
Language: Acceptable — Vivid — Varied Precise — Vigorous — Unified	
Audience Interests and Adaptation: Attention aroused Interest maintained — Knowledge considered Confidence — Beliefs considered	
Ideas: Acceptable purpose — Well developed — Interesting Clear central idea — Accurate — Creative Well supported — Clear — Significant	
Organization: Well introduced — Clear transitions Well divided — Well arranged Well concluded	
General Effectiveness: Mark on a 1 to 5 scale in the box at the right the overall effectiveness of this speech.	

Write additional comments you may have about the speech on the back of this page.

#1 Introduction Rating Scale

Speaker____Date______________

Comments *Score*

Attention _______
Does the introduction attract favorable attention?
Are the attention factors obvious?

Topic _______
Does it state the topic clearly?
It is obvious on which side of the topic the speaker stands?

Lead in _______
Does it lead into the body of the speech?
Are the main points to come clearly stated?
Is it clear where the speaker is headed with the topic?

Credibility _______
Does the intro establish initial credibility?
Does the speaker use the sources of credibility?

Delivery _______
Is the speaker communicative?
Was the eye contact direct?
Were the gestures meaningful?

Total _______

10	7	4	1
superior	average	inadequate	poor

Total Points
50-45 = A
44-40 = B
39-35 = C
34-30 = D
29-00 = F

Comments:

I.D.#________

#2 Speech Rating Scale

Speaker ______________________ Date ______________________

Subject ______________________ Assignment ______________________

Items	Comments	Score
Organization: Clear arrangement of ideas? Introduction, body, conclusion? Was there an identifiable pattern?		
Language: Clear, accurate, varied, vivid? Appropriate standard of usage? In conversational mode? Were unfamiliar terms defined?		
Material: Specific, valid, relevant, sufficient, interesting? Properly distributed? Adapted to audience? Personal credibility? Use of evidence?		
Delivery: Natural, at ease, communicative, direct? Eye contact? Aware of audience reaction to speech? Do gestures match voice and language?		
Analysis: Was the speech adapted to the audience? Did the main points support the purpose?		
Voice: Varied or monotonous in pitch, intensity, volume, rate, quality? Expressive of logical or emotional meanings?		

Total________

10	7	4	1
superior	average	inadequate	poor

I.D.#__________

#3 Consolidated Speech Rating Scale

	1	2	3	4	5	6
Name						
Organization						
Language						
Material						
Delivery						
Analysis						
Voice						
Total						

Evaluator I.D.#__________________

#4 Graphic Scale

Directions Place your I.D. Number here______________

1. In the following pages there are four rating sheets, one for each of four traits of communication.
2. Each of the rating sheets has four vertical lines which represent four speakers.
3. Rate each speaker, according to the descriptions which you feel most nearly apply, by placing an X on the speaker's vertical line in a position corresponding to the description.
4. Use all four rating sheets for each speaker.
5. Use only one line per speaker per sheet--that is, rate speaker 1 on line 1 on all four traits, rate speaker 2 on line 2 on all four traits, etc.
6. These ratings should reflect your honest judgment of the communicative act.

1	2	3	4	IDEAS
				Ideas significant to the audience; ideas well adapted to the audience; assertions well supported; creative treatment of ideas; well analyzed on all points.
				Ideas significant to most of the audience; mostly well adapted to the audience; most assertions supported; mostly creative treatment of ideas; sound analysis on most points.
				Ideas partially significant to the audience; partial adaptation to the audience; some assertions supported and some not supported; average creativity in treatment of ideas; average analysis of ideas.
				Ideas not very significant to the audience; ideas not very well adapted to the audience; most assertions not supported; little creativity in treatment of ideas; weak analysis of ideas.
				Ideas not significant to the audience; ideas not adapted to the audience; assertions not supported; trite treatment of ideas; very shallow analysis.
				ORGANIZATION
				Arrangement easy to follow; planning of organization very evident.
				Arranged overall but unclear on a few points; planning of organization obvious.
				Some points arranged and some not arranged, but can be followed with some effort; planning of organization obvious, but not in all places.
				Arranged in places, but generally hard to follow; planning of organization was hard to see.
				Unarranged and difficult to follow; planning of organization evidently lacking.
				DELIVERY
				Voice and gestures inappropriate to the ideas of the speaker; adapted to audience and situation; free of distractions in voice and gestures.
				Some inappropriate use of voice and gestures in terms of ideas; predominantly adapted to audience and situation; some detractions evident in voice and gestures, but not to the extent of diverting attention from content.

				Voice and gestures sometimes reflect the speaker's ideas; adaptation to audience and situation evident, but not throughout the speech; voice and gestures sometimes distract from the ideas.
				Voice and gestures inappropriate to the speaker's ideas in many places; voice and gestures, not very well adapted to audience and situation; frequent distractions from content because of voice and gestures.
				Voice and gestures not really appropriate to the ideas; voice and gestures not adapted to audience and situation; distractions from content because of voice and gestures are prevalent.
				LANGUAGE
				Concise, varied, and vivid use of language; well adapted to the level of the audience and the topic.
				Language usage partially concise, varied, and vivid; partially adapted to the level of the audience and the topic.
				Language usage predominantly concise, varied, and vivid; mostly adapted to the level of the audience and the topic with few exceptions.
				Language usage partially concise, varied, and vivid; partially adapted to the level of the audience and the topic.
				Language usage not very concise, mostly varied, and not very vivid; adapted to the level of the audience and the topic only in some places.
				Language usage wordy, the same throughout; dull; not adapted to the level of the audience and the topic.

#5 Technical and Professional Speech Rating Blank

	Score
1. Audience analysis 1. Were the materials adapted to the audience? 2. Was the topic adapted to the audience? 3. Was the purpose suitable in terms of the audience?	
2. Organization 1. Were the ideas organized into a logical format? 2. Were there transitions between parts of the presentation? 3. Was the central idea clear? 4. Were the introduction and conclusion sufficient in terms of the structure and the presentation?	
3. Credibility 1. Did the speaker establish credibility early? 2. Was the speaker aware of the various methods of establishing credibility, as evidenced by the presentation? 3. Did the speaker attempt to maintain credibility throughout the presentation?	
4. Research and knowledge 1. Were the ideas supported adequately by evidence? 2. Was research evident, or was the presentation based upon personal experience and supposition? 3. Were questions handled well? 4. Was there a variety of supporting material?	
5. Delivery 1. Was eye contact maintained throughout the presentation? 2. Was there an attempt at meaningful movement and gestures? 3. Was the delivery appropriate for the presentation?	
6. Overall presentation (every member of the group receives the same grade here) 1. Choice of subject 2. Continuity 3. Accomplishment of purpose 4. Establishment of credibility	

excellent	above average	average	below average	poor
16	12	8	4	1

Group number__________________ Total Score_____________
Speaker____
Evaluator____

#6 Speech Rating Blank

Name____Date___________

Subject of the
Speech______________________________Teacher______________________________

After you listen to each of the speakers, answer the following questions with a yes or a no. These questions will be used to help us decide which areas are important to us when preparing to give a speech, and how we may improve our speeches in the future.

Organization--how the speech is put together or arranged.

1. Could you easily pick out the main ideas of the speech?
2. Did the speech have an introduction?
3. Did the speech have a body?
4. Did the speech have a conclusion or summary?
5. Was the speech developed or put together in a way that made it easy for the audience to understand?

Language--the sentence structure of each speech.

1. Were the explanations clear?
2. Was the language easy to understand?
3. Did the speaker make use of pauses to separate ideas from one another?
4. Were there too many "and's" or "uh's" used?
5. Was it easy to tell where one sentence stopped and the next one began?

Material--what the speech was actually about.

1. Was the subject interesting to you?
2. Was the speech easy for you to understand?
3. Did the speaker seem really to know the subject matter?
4. Did the speaker seem comfortable and at ease while giving the speech?
5. Was there eye contact with the audience?
6. Was the speaker aware of how the audience was reacting to the speech?
7. Did the speaker make good use of gestures and body language?
8. Did the main idea stand out above the other ideas?
9. Were there other ideas less important but still necessary in the development of the speech?

Voice--how the speaker sounded.

1. Was the speaker's voice pleasing to the ear?
2. Was the pitch varied--that is, did it go up and down?
3. Was the speaker loud enough?
4. Did the speaker talk too fast?
5. Did the speaker use good expression?

Total Score_______(add the number of "yes" responses) Grade______

#7 Oral Interpretation Rating Blank

Reader________________________________Assignment____________Total Score_____

Item	Comment	Score
1. Introduction: Captures attention, sets the scene and mood, gives needed background, informative.		
2. Material: Of interest to the audience? Is it adapted to the reader, assignment, occasion, and audience? Proper cutting?		
3. Eye contact: Does the reader try to reach each member of the audience? Is there too much dependence on the manuscript? Is there effective character placement?		
4. Articulation and pronunciation: Is it clear, correct, slurred, muffled? Are there defective sounds? Acceptable standards of pronunciation?		
5. Facial expression: Appropriate, varied, adapted to the reading? Does it aid in expressing the emotions in the selection?		
6. Poise: Confident, at ease, personality pleasing, moves easily, projected to the audience? Aware of audience reaction to the reading?		
7. Bodily action: Is the reader animated? Are posture, action, and gestures constructive or distracting? Are gestures used effectively, varied, suited to content and purpose?		
8. Vocal quality: Is it pleasant to listen to? Is there sufficient variety, projection, clarity? Acceptable volume?		
9. Rate and timing; Are rate and pauses varied and suited to content and purpose? Too fast or too slow?		
10. Content: Communicated author's intent as stated in the introduction? Was the content adequate to support the reader's goal as stated in the introduction?		

10 superior	7 average	4 inadequate	1 poor

Total Score:

Additional Comments:

I.D. #_________

#8 Communication Objects Rating Scale

Name of
Communicator____

1. How well was the self-concept explained?

superior	excellent	above average	average	fair	below average	poor

2. How well was the concent of what others think explained?

superior	excellent	above average	average	fair	below average	poor

3. How well was the ideal concept explained?

superior	excellent	above average	average	fair	below average	poor

4. How well was the undesired concept explained?

superior	excellent	above average	average	fair	below average	poor

5. How well are the differences among the four objects clarified?

superior	excellent	above average	average	fair	below average	poor

6. How well do you think the communicator was prepared?

superior	excellent	above average	average	fair	below average	poor

Name of Evaluator____

(Place other comments on back of sheet.)

#13 Short Speech Feedback Form

Name____Topic______________________

Item	1	2	3	4	Comments
Clearly stated thesis					
Examples Direct support Clear explanation					
Eye contact					
Gestures					
Body position					

Evaluator____Grade____________

Key:
1 = Needs a lot of work.
2 = Doing well some of the time.
3 = Doing well most of the time.
4 = Excellent job--keep it up.

Appendix F

Interview Rating Form (from Seibold & Meyers 1985)

Interview Rating Form

Please circle the number under each item which most closely reflects your assessment of the student's performance in that area.

1. Initial impressions of interviewer (first minute)
    favorable 1 2 3 4 5 unfavorable
2. Opening comments
    appropriate 1 2 3 4 5 inappropriate
3. Physical presentation
    appropriate 1 2 3 4 5 inappropriate
4. Composure
    excellent 1 2 3 4 5 lacking
5. How clearly interviewer expressed purpose of interview
    clear 1 2 3 4 5 unclear
6. Nonverbal communication (eye contact, posture, enthusiasm)
    excellent 1 2 3 4 5 lacking
7. Verbal communication (voice, articulation, fluency, volume, vocabulary)
    excellent 1 2 3 4 5 needs work
8. How well interviewer listened to your responses
    excellent 1 2 3 4 5 needs work
9. Providing feedback to interviewee
    excellent 1 2 3 4 5 lacking
10. Interview technique
    favorable 1 2 3 4 5 unfavorable
11. Transitions between question areas
    clear/ appropriate 1 2 3 4 5 unclear/ inappropriate
12. Ability to keep interview session moving
    excellent 1 2 3 4 5 lacking
13. Questions asked
    appropriate 1 2 3 4 5 inappropriate
14. Clear 1 2 3 4 5 Unclear
15. Ability to use interviewee's responses in later questions
    favorable 1 2 3 4 5 unfavorable
16. Control over interview
    excellent 1 2 3 4 5 lacking
17. Student preparation
    excellent 1 2 3 4 5 lacking
18. Overall assessment of interviewer's performance
    excellent 1 2 3 4 5 poor
19. Additional comments which may be of help to the student

---

20. Would you be willing to participate in another interview in the future? YES NO (Please Circle)

Name________________________________ Title____________________

Address____

Phone____________________

Appendix G

Rating Form for Conversational Skills (from Spitzberg & Hurt 1987)

Conversational Skills Rating Scale

Person Being Rated: ______________________________________

Person Doing Rating: (ID#__________________________(Male or Female) Date:______________

Rate the conversant according to how skillfully he or she used, or didn't use, the following communicative behaviors in the conversation, from:

1	INADEQUATE (use was awkward, disruptive, or resulted in a negative impression of communicative skills)
3	ADEQUATE (use was sufficient but neither very noticeable nor excellent. Produced neither positive nor negative impression)
5	EXCELLENT (use was smooth, controlled, and resulted in positive impression of communicative skills)

Circle the single best response for each behavior:

1 2 3 4 5 (1) Use of eye contact
1 2 3 4 5 (2) Initiation of new topics
1 2 3 4 5 (3) Maintenance of topics and follow-up comments
1 2 3 4 5 (4) Use of time speaking relative to partner
1 2 3 4 5 (5) Interruption of partner
1 2 3 4 5 (6) Speaking rate (neither too slow nor too fast)
1 2 3 4 5 (7) Speaking fluency (avoided pauses, silences, "uh," etc.)
1 2 3 4 5 (8) Vocal confidence (neither terse nor nervous sounding)
1 2 3 4 5 (9) Fidgeting or playing with things (e.g., pencil, rings, hair)
1 2 3 4 5 (10) Shaking or nervous twitches (weren't noticeable)
1 2 3 4 5 (11) Posture (neither too closed/formal nor too open/informal)
1 2 3 4 5 (12) Unmotivated movements (eg, avoided swaying feet, tapping fingers)
1 2 3 4 5 (13) Asking of questions
1 2 3 4 5 (14) Nodding of head in response to partners
1 2 3 4 5 (15) Lean toward partner (neither too far forward nor too far back)
1 2 3 4 5 (16) Speaking about partner (involved partner as a topic of conversation)
1 2 3 4 5 (17) Speaking about self (didn't talk too much about self or own interests)
1 2 3 4 5 (18) Encouragements or agreements (encouraged partner to talk)
1 2 3 4 5 (19) Use of humor and/or stories
1 2 3 4 5 (20) Vocal variety (avoided monotone voice)
1 2 3 4 5 (21) Vocal volume (neither too loud nor too soft)
1 2 3 4 5 (22) Expression of personal opinions (neither too passive nor too aggressive)
1 2 3 4 5 (23) Facial expressiveness (neither blank nor exaggerated)
1 2 3 4 5 (24) Use of gestures to emphasize what was being said
1 2 3 4 5 (25) Smiling and/or laughing

For the next five items, rate the person's overall conversational performance:

(26) UNSKILLFUL MANAGEMENT: 1 2 3 4 5 :SKILLFUL MANAGEMENT
(27) INEXPRESSIVE: 1 2 3 4 5 :EXPRESSIVE
(28) INATTENTIVE & UNRESPONSIVE: 1 2 3 4 5 : ATTENTIVE & RESPONSIVE
(29) ANXIOUS & NERVOUS: 1 2 3 4 5 :RELAXED & CONFIDENT
(3O) INAPPROPRIATE & INEFFECTIVE: 1 2 3 4 5 :APPROPRIATE & EFFECTIVE

NO GURU, NO METHOD, NO TEACHER: THE COMMUNICATION DOMAIN AND THE NACSL

by Stephen P. Witte

Adapting an earlier expression to a new context, some wag once quipped that a camel might be defined as a horse designed by a group of experts. Such blanket cynicism about expert groups and their efforts is probably unwarranted given the fact that much of what gets done in this world, and sometimes gets done fairly well, quite obviously depends on group efforts. Nevertheless, there is always the possibility that compromise and consensus, outcomes of the very activities that often permit groups to arrive at closure with respect to a task at hand, will lead to the creation of a camel when a horse was needed and perhaps intended. The probability that a group will construct a camel instead of a horse, it seems to me, increases considerably under three conditions: (a) when the task at hand demands accommodating multiple disciplinary perspectives and special interests that may not be similarly understood by all the parties involved, (b) when the knowledge that a whole is usually greater than the sum of its parts is in danger of being swallowed up by a narrow focus on isolated components, and (c) when history provides neither a model nor a prototype for what the group must produce. These are, I believe, at least some of the conditions under which the current effort to develop a National Assessment of College Student Learning (NACSL) operates, both with regard to the assessment in general and with regard to what has been called "the communication domain" in particular. Without intending or implying any disrespect toward cultures in which the dromedary figures importantly and without intending or implying any disrespect for colleagues who, like me, are or have been group participants in the development effort, I worry a good bit that in the end the NACSL will have much more in common with a camel than it will with a horse, even though most of us involved in this effort probably carry about an image of a horse in our minds.

This worry of mine is greatly exacerbated by my own forays (Witte, Trachsel & Walters, 1986; Witte, Cherry, Meyer & Trachsel, *in preparation*) into the history of large-scale assessments of, particularly, writing. That history, both early and late, gives some point to the waggish quip about expert-group processes and the sorts of camels that those processes often engender. It was my sense of the place that the current NACSL effort might ultimately find in assessment history that prompted me to borrow part of my title for the present paper from a Van Morrison (1986) album: From the perspective of one who has spent a good bit of his career trying to understand how people produce and use "texts" to get things done, Morrison's title serves for me both as a statement of what assessment history seems capable of contributing to the development effort and as a summation of the work to date in planning a communication component for the NACSL (for overviews of that work, see Greenwood, 1992, and NCES, 1992). Because history seems to me incapable of providing a model for the "communication component" of the NACSL, I was also prompted to appropriate Morrison's title in order to call attention to the formidable problems that lie ahead with regard to conceptualizing the NACSL, both with regard to an underlying theory of language (and thinking) and with regard to how the assessment itself might be carried out.

Within the context of the consensus building (cf. Greenwood, 1992; NCES, 1992) that informs the current NACSL development effort, whether what I have to say will be seen as proffering anything by way of portable substance or as pointing out agreeable directions are for me open questions. My hope, of course, is that this paper will contribute somewhat more to the design of a horse that looks and functions as horse than it contributes to the construction of a camel or some other as yet undefined beast such as a *corse* or a *hamel*. My own thinking about the NACSL generally and about its communication component particularly is informed by the assumption that what is and can be assessed is affected significantly by how assessment is executed or carried out. Accordingly, I've organized the following remarks into two major sections and a brief and concluding third section. The first section focuses on "the communication domain," and the second proposes an approach to assessment that might hold some promise for tapping into the domain without substantially altering its underlying nature or structure in the process.

The Communication Domain: Writing and Reading

Section I of this paper attempts to respond to two of the three directives given to me in writing and, subsequently, reinforced during discussions in Washington with NCES personnel the last week in September of 1992. Those two directives are the following:

(1) To "address and make recommendations in regard to . . . a preliminary listing of selected [reading and writing] skills and a rationale for [their] inclusion" among the "skills" to be assessed through the NACSL (Corrallo, 1992, p. 5), and

(2) To "address and make recommendations in regard to . . . levels of achievement or proficiency for each of the skills in the preliminary listing" (Corrallo, 1992, p. 5).

These two directives might appear at first glance to some as ones that can be easily dispensed with, as they seem to have been traditionally in most, if not all, large-scale assessment efforts. Yet as both Venezky (1991) and White (1991) suggest in the papers they prepared for the November 1991, NCES Workshop on Assessing Higher Order Thinking and Communication Skills of College Graduates, there are a host of mutually informing conceptual and practical issues embedded in the two directives, issues that need to be exposed or unpacked in advance of any "listing of selected skills," however "preliminary" that listing may be. Accordingly, the present section is divided into four parts--the first dealing with some practically situated theoretical issues and concerns, the second providing an example that elaborates and supports the theoretical position on communication endorsed in the first part, a third proffering "a preliminary listing of selected skills" that is grounded in what can be gleaned from extant research about how writing and reading gets done and functions in different settings, and a fourth part that addresses the question of "levels of achievement or proficiency."

Some Theoretical Issues Practically Situated

Any "preliminary listing of selected [reading and writing] skills" that might become a part of the NACSL would, it seems to me, embed certain assumptions about the nature of communication. Thus, in advance of a "preliminary listing," I want to set out briefly how I think communication ought to be conceptualized. And I'll attempt to do that through a comparison. At the risk of oversimplifying some fairly complex issues, I want to suggest that communication can be viewed in at least two ways, one that can be easily understood but that is fundamentally out of touch with the way people actually use symbols to communicate with others and a second that is more difficult to get a handle on but that reflects communication practice far better.

The first of these may be called the conduit or ***transmission view of communication***. Informing this view is a metaphor borrowed from early work in radio: A transmitter sends a signal to a receiver, which in turn interprets the signal as it was transmitted, unless, of course, the signal is distorted. Appropriated to human communication studies (for a survey and a critical appraisal of this literature, see Bickard, 1980), the "transmission" metaphor and what it implies about the communication signal and the mechanical sorts of process by which the signal is sent and received has led to at least three misunderstandings. One of these misunderstandings is that "meaning" is a property of (i.e., "resides in") texts, whether oral utterances or written statements. A second misunderstanding is that the meaning presumably contained in the encoded signal (the "text") is simply duplicated or replicated in the mind of the decoder (i.e., reader or listener) in much the same way as a text file is duplicated from a hard disk to a floppy. According to the transmission view of communication, writers or speakers put meaning into "texts," and readers or listeners simply take it out again. A third, and not unrelated, misunderstanding is that "texts" are autonomous, that they can stand alone on their own, independent of the contexts in which they might have been produced or used. According to the transmission view, texts are a lot like boxes or tin cans: It doesn't matter where the boxes or the tins might be taken or who might happen to open them and under what circumstances, their contents remain the same, invariable across space and time and context, only to be processed by a reader in much the same way as a radio receiver processes the signal it receives. According to this view, meaning exists as a thing or an object, and the linguistic code serves merely as the package in which meaning is delivered to an individual recipient who has merely to unwrap the package and find the meaning therein.

But meaning does not reside in texts as the transmission model of communication assumes. Even seemingly simple, straightforward texts such as "**FIRE!**" are not necessarily transparent. If that text is uttered by someone running down a hallway past doors from which smoke pours, "**FIRE!**" could "mean" one thing; but if, on the other hand, it is uttered in an equally loud voice by sergeant in a battlefield foxhole, "**FIRE!**" could "mean" something entirely different.

And if it were shouted loudly by one participant in a meeting where all participants are seated around a conference table in a room filled with bright sunshine and with no visible or olfactory evidence of fire, "**FIRE**!" could "mean" for the other participants any number of things, some of which would not likely flatter the speaker. The point is that, as Malinowski (1946) showed, meaning is always context-dependent or, to follow Fillmore's (1977) convincing demonstrations, meaning is always "relativized to scenes" (p. 84): Neither individual words nor combinations of words **contain** meaning.

It is the context-dependent, relativistic nature of meaning (and meaning construction) that is conspicuously absent from the transmission view of communication. Nevertheless, the transmission model of communication is, it seems to me, the dominant popular view, popular in the sense of its being widely held. Its tenets, I would argue, are so much a part of the fabric of everyday representations of both "texts" and of reading and writing (as well as of speaking and listening) processes that the view might even be regarded as one of the "metaphors we live by" (cf. Lakoff & Johnson, 1980), even if the metaphor is at odds with reality in the same way that the everyday words "sunrise" and "sunset" are at odds with a post-Copernican view of the universe.

Nevertheless, the salience for some of the transmission view of communication can hardly be denied. It shows up, for example, in the blanket injunctions of some textbooks writers and composition teachers against constructions such as passives, nominalizations, and subordinate clauses on the grounds that such linguistic structures presumably obscure the meaning they contain. It informs in important ways the results garnered from surveys of writing in workplace settings (e.g., Barnum & Fischer, 1984; Bataille, 1982; Faigley, Miller, Meyer & Witte, 1981; SCANS, 1992; Stine & Skarzenski, 1979; Storms, 1983), results that ascribe great importance to descriptors such as "clarity" or "conciseness," which are represented as universal qualities of "good writing" because they are presumed to make the meaning contained in, or packaged by, the language transparent, more readily accessible. The transmission view--by virtue of its emphasis on language as a container of meaning--also shows up in claims that a written text permits direct access to the thinking that informs it, an assumption that Horace Mann invoked in arguing that the 19th-Century oral examination system of the Boston schools be replaced by a written one: For Mann, a written text functions as "a transcript, a sort of Daguerreotype likeness of the state and condition of the pupils' minds, [which] is taken and carried away, for general inspection" (Mann, 1845, p. 334). Mann's notion of a written text as a "picture" of thought and thinking implies a one-to-one correspondence between the writer's thoughts and words and, further, a one-to-one correspondence between the writer's words and the reader's thoughts. Mann's position, it should be pointed out, is not too far removed from the position White (1991) takes in proposing that portfolios of writing be used to assess higher order thinking, a position that assumes--quite improperly, I think--that (a) for higher order thinking to have occurred, it must have manifested itself in printlinguistic code and (b) that "poor" prose (however defined) is necessarily a sign of "poor" thinking. Although one might be able to make the case for inferring higher order thinking from "good" prose texts, it is--in my view--no more possible to infer the absence of higher order thinking from "poor" writing than it is to infer the lifestyle of a pig from a sausage patty. In addition, the transmission view of communication--particularly when it is used to privilege printlinguistic texts--also shows up in the many ways in which Western societies equate the word "illiterate" with terms such as a "ignorant" and "stupid" (cf. Stubbs, 1980, p. 50; Oxenham, 1980, pp. 42-57); and it appears in equations, such as Farrell's (1983), that forge an identity between standard English usage and intelligence.

The transmission view of communication shows up in many standard assessment practices as well, practices that may reveal some things about one's command of some "basic" dimensions of communication but that reveal very little about one's ability to communicate effectively in particular situations. The transmission view certainly informs those multiple-choice tests--called "indirect measures" in the literature on writing assessment (e.g., Breland & Gaynor, 1979)--that purport to assess writing or reading independently of any naturally occurring context of language production or use. But the view also shows up in so-called "direct" assessments of writing. In such assessments, writers write for no other purpose than to accommodate someone else's need for an evaluation, and their "texts" are often judged by readers who read for no other reason than to judge. As far as I know, such purposes for writing and reading do not obtain in any naturally occurring context for communication. But, as Charney (1984) has pointed out with regard to "holistic" assessments of writing, the most widely practiced form of direct assessment, the scores and the inferences drawn from them are assumed to be "valid" simply because of the presence of the writing sample, a text that is assumed to "mean" independently of any "authentic" communicative context for production or use.

The fact that "assessment texts" are written and read independently of any naturally occurring purpose or context for writing or reading strongly suggests that those texts are regarded as autonomous artifacts. Viewing texts as autonomous artifacts--as self-contained vehicles for or containers of meaning--is, it should be remarked, largely consistent with the principles of "New Criticism" (for a brief history and an outline of major tenets, see Willingham, 1989) that dominated English departments when modern holistic scoring began taking shape with the "general impression" method used by Godshalk, Swineford, and Coffman (1966). Whether such "assessment texts" are the stable, meaningful artifacts they are assumed to be is, however, open to question because those texts cannot be rated reliably without the use of rigorous rater-training procedures (e.g., Diederich, 1974; White, 1985), which may not even be entirely replicable across assessment contexts (cf. Freedman & Calfee, 1983). Such training procedures, it might reasonably be urged, do as much to assure that readers come to agree not to disagree about features of texts that authorize variant readings and judgments as those procedures do to insure that readers agree about what constitutes "good" writing. In the wake of what reading research suggests about the role of readers' prior knowledge in their understandings of texts (for overviews of this literature, see Alexander, Schallert & Hare, 1991; Schallert, 1982) and in the wake of what, for example, reader-response criticism (for an overview, see Rabinowitz, 1989) suggests about the nature of interpretive reading, it is probably worth inquiring whether the generally low interrater reliabilities that obtain--in the absence of rigorous training--for holistic scoring have more to do with multiple "valid" readings of any given text than with the random error of measurement that reliability formulas are designed to indicate.

But in naturally occurring situations people don't communicate ***to*** one another, as the transmission model implies they do; rather they communicate ***with*** one another. The differences signaled by the two prepositional markers are, I think, extremely important ones. First of all, the former invokes the image of language and text as a "container" of meaning; and the latter invokes the image of language and text as a delimited set of symbols that defines an important, but a single, component of the socially situated field on which participants meet in meaning-constructive activity, the outcome of which is never determined--as the example(s) of the "**FIRE!**" text(s) illustrate--on the basis of text alone. Second, the former ***communicate to*** implies a language that functions to reify the line that divides the "self" from the "other," and ***communicate with*** implies a language that functions to unite the "self" and the "other." Third, ***communicate to*** implies that miscommunication is solely a function of text, and ***communicate with*** implies that miscommunication is just as likely to be function of text in context as, again, the "**FIRE!**" texts suggest. The transmission view of communication denies this *in situ* nature of texts, communication, and communicators.

In contrast to the view of communication that sees texts as containers for meaning, that sees "textual meaning" as stable across contexts, that sees writing as packaging thought and meaning in words, and that sees reading as taking thought and meaning out of words is a very different view. I shall refer to this second view as the ***constructivist view of communication***, of reading and writing (as well as speaking and listening). Even though some scholars in writing such as Bazerman (1983), Bizzell (1982a, 1982b), or Faigley (1985, 1986) might prefer the term *social* over the term *constructivist*, and others such as Nystrand (1986, 1989; Nystrand, Greene & Weimelt, *in press*) might prefer the terms *dialogic* or *social-interactive*, and still others (e.g., Berkenkotter, 1991; Flower, Stein, Ackerman, Kantz, McCormick & Peck, 1990; Geisler, 1991; Greene, 1990, 1991) might prefer the term *sociocognitive*, the term *constructivist* would probably find some support from students of writing (e.g., Berkenkotter, Huckin & Ackerman, 1991; Bracewell & Breuleux, *in press*; Brandt, 1986, 1992; Brodkey, 1988; Flower, 1989, *in press*; Flower & Hayes, 1984; Myers, 1985, 1990), of reading (e.g., Bazerman, 1985; Beers, 1987; Spiro, 1980; Spivey, 1987; Tierney & Pearson, 1983; Tierney, LaZansky, Raphael & Cohen, 1987), and of reading and writing (e.g., Bracewell, Fredriksen & Fredriksen, 1982; Flower, 1989; Flower, Schriver, Carey, Haas & Hayes, 1992; Spivey, 1984, 1990).

Some readers might, of course, object that I've constructed out of the term *constructivist* a sort of Procrustean bed into which I've rather unceremoniously shoved a number of distinguishable theoretical positions; and other readers might suspect that I've created a Procrustean bed in order to assimilate into my own "constructivist" perspective (cf. Witte, 1991a, 1991b, 1991c, 1991d, 1992; Witte, Nakadate & Cherry, 1992) positions that have a great deal of theoretical integrity in their own right and that I have, thereby, improperly and unconscionably reduced some important theoretical differences to a matter of "mere semantics." However, it is not my intention to deny or to make light of the differences between others' perspectives and my own;

nor is it my intention to deny the differences that exist among the perspectives of others. Moreover, I recognize that some of the work I've cited addresses issues and problems that some other work I've cited does not. However, this latter is a fact which--in my judgment--has little or nothing to do with theoretical compatibility. In any case, my concern here is primarily (although not exclusively, as will become clear) with theoretical commonalities rather than with differences among the positions I've "packed into" the term *constructivist*. And I don't feel compelled to apologize for liking better than some other term a term that I've used before and that others found to be useful long before I did.

The common grounds I see among the various positions I've grouped together as representing the *constructivist view of communication* distinguish it from the *transmission view of communication*. According to the constructivist view, meaning does not reside in texts even though writers may manipulate features of the texts they write (or speak). Rather than containing meaning, texts function as organized sets of signs of what writers intend, signs of meanings that writers (and speakers) represent to themselves. And because words and larger collections of "text signs" can and do have multiple "interpretants" (cf. Witte, 1992), writers themselves often construct new meanings as they write, monitoring (in a dialogic fashion) their own acts of writing by reading the visible signs they have employed, sometimes constructing those new meanings through the symbolizing process itself and sometimes constructing those new meanings as they evaluate their intentions against the meaning they construct for the text under way. As organized sets of signs, texts cue meaning-constructive processes in readers (and listeners) when there exists a situational demand for meaning-constructive activities, a demand that writers can sometimes (and sometimes not) anticipate and accommodate through the particular signs and collections of signs they employ. While it is true that the "text sign" affects the meaning a reader constructs for a given text, that meaning is also conditioned or constrained by what the reader brings to the text by way of cultural background, prior knowledge, purpose for reading, and the context in which the reader's meaning-constructive acts occur. This latter point leads to another tenet of the constructivist view, namely, that texts are not autonomous. Whether viewed from the perspective of the writer or the reader, texts are context situated, both in terms of the context of production and the context of use, and the meanings constructed for them are no less situated in context: When a text moves from its context of production to a context of use or from one context of use to another, there is always the possibility that the meanings constructed for it will differ. According to the constructivist view of communication, meaning is not merely transmitted through the vehicle of language as though down a one-way street from the mind of a writer to the mind of a reader. The whole process is much more interactive, much more dialogic, much more constructive, as writer and reader meet at the juncture of the "text sign."

To some, it may appear that the meaning-constructive processes of both writers and readers are exclusively individual and decidedly cognitive. However, such is not the case. In fact, the cognitive and the social complement one another. Even though it involves cognition, the intersection of reader and writer at the "text sign" is a decidedly social function, just as language itself is a decidedly social phenomenon. The meaning-constructive processes of both writers and readers are also collaborative and social, dialogic and interactive, although less obviously so in the case of printlinguistic texts than oral ones. In addition to the obvious facts that the processes of reading and writing are always situated in particular social contexts and that the meanings constructed are constrained by what meanings are possible within and supported by those contexts, readers and writers collaborate with other readers and writers by virtue of the fact that any new text is always in some sense a response to at least one other text, which is itself always a response to at least one other text, and so on. Accordingly, every new text also enters a stream of other discourses--some of which may be known only to the writer, some of which may be known only to the reader, some of which may be known to both the writer and the reader, and some of which may be known to neither the writer nor the reader. And in many contexts--whether in homes, schools, or workplaces--this collaborative and social dimension of reading and writing becomes much more overt when the meaning-constructive processes associated both with texts in progress and texts being read are conditioned by, for example, conversations with other people. And although it would be foolish to argue that cognition does not figure in such social events, it would be equally foolish to argue that such acts do not depend on social processes.

In either case, the position of the present essay is that the verb *to communicate* signals much more by way of the cognitive and social complexity of communicative acts than is allowed by the transmission model. All language acts entail both the cognitive and the social. To construct meaning, whether as a reader or a writer, involves the use of language, in many of the social ways I've indicated, as a tool for learning. It is, of course, common knowledge that readers learn as they read; it is less widely known, though nonetheless obvious, that writers learn as they write, constructing what for them are new meanings and new relations with the world through symbolization processes. Engaging in symbolization processes not only appears to mediate all learning (cf. Sebeok, 1991; Vygotsky, 1934/1986) but would also appear (given people's memories of communicative events) to insure that learning of some kind occurs whenever one engages in symbolization. Moreover, engaging in symbolization processes such as reading and writing appears to force one to reflect, sometimes consciously and sometimes not, on both what and how one knows as he or she judges the potential meaning signaled by a given text, whether one being read or one being written, against his or her internal representation of the world. The learning and reflecting functions, both of which involve "critical thinking" in the construction of meaning through symbolization processes, point to a third function embedded in communicative acts, namely, problem solving. The importance of this third function becomes apparent when we try to represent occasions or purposes for communication as **not** involving one or more problems to be solved. Problem solving itself is a meaning-constructive process that entails problem representation, which involves reflection on what one knows, and that entails the use of strategies for orchestrating the processes, the topics, and the signs of one or more symbol systems to construct and represent a solution to the problem. In short, the verb ***to communicate*** points to more, and to more complex, cognitive and social processes than the transmission model allows.

As I have intimated, these cognitive and social dimensions of the constructivist view of communication can and should be extended to an understanding of the roles nonlinguistic symbol systems play in writing and reading, a position I urged in a recent article on the basis of a series of case studies of writers working in contexts from home to school to workplace (Witte, 1992). In that article, I argued--among other things--that writers and readers not only construct meaning from the sets of linguistic signs that are traditionally called "texts," but that they also do so for organized sets of signs that are not usually considered "texts" by people who study reading and writing. Among these nonlinguistic "texts" are a host of "social texts" (e.g., the way in which a family seats itself at a dinner table, the physical arrangements of classrooms, the nonverbal behaviors of people in a meeting) for which writers must construct meaning even during the process of producing more traditional printlinguistic texts. In addition, both writers in the process of producing traditional texts and readers in the process of reading traditional texts must often draw on the meanings they construct for "texts" (such as graphs and photographs) that depend heavily and sometimes exclusively on nonlinguistic symbol systems or for "texts" (e.g., television newscasts, scholarly articles, engineering reports, billboard and newspaper advertisements, telephone book Yellow Pages, newspaper stories) that routinely incorporate both linguistic and nonlinguistic symbol systems. The fact that "texts" which depend on nonlinguistic symbol systems probably constitute a majority of the non-oral "texts" people encounter and use on a day-to-day basis is testimony to their importance in contemporary culture and points to the necessity of accommodating them in current theories and studies of reading and writing. Moreover, the fact that many people in workplace settings not only have to use and produce (often in highly collaborative ways) such "texts" on a routine basis strongly suggests the critical role they actually play in the meaning-constructive activities that have traditionally been associated only with writing and reading printlinguistic texts. The production and use of nonlinguistic or partially nonlinguistic texts made possible by the computer technology presently available (and likely to be enhanced and to become more widely available in the near future) to writers and readers in homes, schools, and workplaces suggest an even more pressing need to incorporate the production and use of "texts" relying on nonlinguistic symbol systems into our understandings of writing and reading. In concert, the observations (some would say, speculations) I have offered--in the context of the constructivist view of communication that finds, I believe, strong support in the literature on reading and writing--about the role of both "social" and "nonlinguistic" texts in meaning-making activities strongly suggest that higher order thinking probably manifests itself in nonverbal ways that are not necessarily accorded much status (cf. Gardner, 1983, 1985), and those observations suggest that higher order thinking, which is traditionally associated with primarily verbal texts, might actually be mediated cognitively and socially through the use of nonlinguistic and historically prior symbol systems (cf. Sebeok, 1991; Witte, 1991a, 1992).

Practicing Theory, Theorizing Practice: An Example

Theories--such as the respective theories embedded in the *transmission* and *constructivist* views of communication I have endeavored to characterize briefly--are useless unless they help us to understand known practices or performances and to predict the nature of future practices or performances. This is particularly true when it comes to the tricky business of assessing communication "skills": What we understand those "skills" to be and how we go about assessing them depends on our underlying theory of communication. The two perspectives on communication I have outlined, both the one I have rejected and the one I have endorsed, need to be tested against actual practice. To a substantial degree, they have already been, as evidenced in the literature on reading and writing I have cited and in a large number of other studies as well. But in the context of the development of a communication component for a National Assessment of College Student Learning, the relative value attributed the two respective views probably ought not to be determined solely on the basis of the fairly abstract and general arguments I have advanced, however well those arguments might seem to accord with the experiences and understandings of others. More is required. What is required is a grounding or demonstration of the principal claims being made here.

Interestingly (and appropriately) enough, what can be gleaned of the NACSL development effort to date appears to support the rejection of the *transmission* view and to support the acceptance of the *constructivist* view of communication I have outlined. Let me try to explain.

The NACSL development effort--of which the present paper, its author, and its initial readers are a part--appears to have been inspired in some important ways by Goal 5, Objective 5, of the National Education Goals Panel's (NEGP) *America 2000* (1991). Goal 5 reads, "By the year 2000, every adult American will be literate and possess the knowledge and skills necessary in a global economy and exercise the rights and responsibilities of citizenship" (NEGP, 1991, p. 39). Objective 5, which is subsumed under this seemingly rather ambitious "social" goal, pertains specifically to college student learning and has apparently been interpreted by government officials as authorizing the NACSL. Objective 5 reads, "The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially" (NEGP, 1991, p. 40).

Having read Goal 5 and Objective 5, try now making sense of Objective 5 as though it were an autonomous text (i.e., one that stands alone as it does in much of the printed literature associated with the NACSL effort), as though it "contained" a meaning that was transparent, and as though it provided a "picture" of underlying (higher order) thinking. Doing so is, of course, impossible because in order to make any sense of the statement, we must "inform" it by linking what its text signs signify for us to our own prior knowledge of concepts such as "advanced," "ability," "think," etc., by linking those signs to our knowledge of what such terms represent in our own experience, and by linking those signs to our knowledge of how we constructed meaning for such symbols when they appeared in previous texts we have encountered. Moreover, it is impossible to make sense of Objective 5 according to the tenets of the transmission view of communication because the processes by which we as readers "inform" the text signs individually and the text sign as a whole would have to be situated within a particular context for reading, which would include our purpose for reading the text in the first place. In addition, because not all of us readers have similar backgrounds, different ones of us would predictably construct different meanings for key words or phrases such as "advanced ability." Indeed, most of the papers presented and discussions occurring during the November 1991, NACSL Study Design Workshop (see the overviews and summaries in Greenwood, 1992, and NCES, 1992) provide ample evidence that whatever terms such as "advanced ability" come to mean within the NACSL effort, those meanings will be constructed collaboratively over time by many individuals.

When I first read the text several months ago--well before I was asked to write this paper--in (I think) *The New York Times*, I read it very hurriedly, not even pausing to consider how exactly it might mean for me, and then I promptly forgot I'd ever read the thing. It was only after agreeing to write this paper and seeing Objective 5 reproduced in documents sent to me by NCES that I remembered even having read the statement before. But since that time, I have puzzled for many hours over the sort of meaning I might or should construct for the statement and, hence, "inform" it with. And to a large extent, my writing the present paper has been a protracted process during which I've tried to construct meaning for that statement. And partly because the text affords the reader little help in interpreting or constructing meaning for the key single terms (such as "proportion," "advanced," "communication," and "substantially") and key phrases (such as "college graduates," "communicate effectively"), the statement admits multiple readings, multiple constructions. And of those possible readings, at least one does nothing to flatter the Goals Panel, as the following account suggests.

Coming as Objective 5 does from the NEGP (no doubt a body of thoughtful, respected, and well-intentioned individuals toward whom we ought to show a great deal of respect), we might regard the statement as signalling something very thoughtful. But as we read the statement, we might find ourselves a bit troubled by it, not so much because of what we might think it says and means explicitly but because of the inferences we can construct from it. For example, as we read the text, we might think that it makes some rather broad-sweeping but implicit claims about "educational problems" (i.e., that a significant number of college graduates don't possess certain "advanced" abilities) and, further, that--on the basis of that claim and without pausing to pick up a warrant or two and some corroborative evidence--it then sets out an educational goal for the future (i.e., "The proportion of college graduates . . . will [read, *needs to*] increase substantially"), as though the claim (i.e., the "problem") could simultaneously function as its own warrant and its own supporting data. To be sure, some of the troubling thoughts we might have about the claim would be calmed or soothed considerably if a body of evidence were marshalled (or cited) in support of the claim being advanced. Yet there appears in the statement nothing that either warrants or supports the claim.

But since we are social beings who exist in a culture that typically expects its peoples to show some measure of respect to its leaders, we might be altogether willing to accept the claim as factually grounded and the stated goal as necessary. But, then, we might happen to read a portion of the *Interim Report* issued by the National Education Goals Resource Group in March 1991--after the initial 1990 publication and distribution of *America 2000*--in which appears the statement that "neither national nor state information is currently available on the ability of college graduates to 'think critically, communicate effectively, and solve problems'" (quoted from NCES, 1992, p. 3), thereby raising once again the questions of whether the implicit claim of Objective 5 were grounded at all and whether the consequent goal were justified. Thus, situated alongside the *Interim Report*, the NEGP's Objective 5--which did not change between the 1990 printing of *America 2000* and the April 1991, revision--might permit the more cynical among us to construct the inference that a lot more is being "communicated" than was probably intended about the ability of at least some people to "think critically, communicate effectively, and solve problems"; and it might be urged, accordingly, that embedded within the text sign as a whole is very little that permits the reader to infer the credibility and the ethical appeal that one might hope to infer from a national statement of policy issued by such a group of people and endorsed by a U.S. President.

It might, of course, be objected that this unflattering reading of Objective 5 is the result of my having constructed meaning for the text "out of context." But out of what context? The context that would include the Panel's deliberations and the sorts of evidence it considered in drafting Objective 5? Perhaps. And I would certainly find interesting indeed the transcripts of those deliberations, copies of correspondence received and produced by Panel members, and the trail of evidence the Panel followed in arriving at its implicit claim. But, in point of fact, that was not the context in which I've been reading Objective 5 of late, nor is it the one authorized by NCES through either its directives to authors or the printed materials sent to them. In addition to texts such as memoranda incorporating the directives to the authors, NCES reports documenting educational trends, portions of reports such as the Department of Labor's *Learning a Living* (SCANS, 1992), authors were sent Greenwood's (1992) comprehensive summary of the November 1991, NACSL Study Design Workshop, a paperbound report of group discussions during the Workshop (NCES, 1992), about two dozen papers commissioned for the Workshop, and about as many (and perhaps more) written responses to those commissioned papers. Not among any of the printed materials I received from NCES was a copy of *America 2000* or any background material pertaining to its production. However, if I remember correctly, all of the 1991 commissioned papers and the two summary documents from NCES (i.e., Greenwood, 1992; NCES, 1992) did quote Goal 5, Objective 5, as I have done in the present paper. But of all the documents sent by NCES, only one, a commissioned paper by Dunbar (1991), actually included a bibliographical entry for *America 2000*, the source document for Goal 5, Objective 5. Moreover, even though some of the writers of the 1991 commissioned papers seemed genuinely skeptical about the value or the feasibility of a National Assessment of College Student Learning, it is important to note that none of them ever wrote or stood up in meetings and said "NOT!"

Now, given the unflattering interpretation I hypothesized for Objective 5--a construction of meaning that the text sign permits and a construction that doesn't require a great deal of mental work--and given the absence of *America 2000* from the materials I received and the general absence of bibliographical citations for a document quoted across the 1991 commissioned papers, we can begin to see how context affects both meaning construction and text value. In these regards, it should be pointed out that the above construction of meaning for Objective 5 showed up in none of the 1991 commissioned papers and, as far as I can tell, in none of the small-group oral discussions of the papers. The question is "Why not?" The reason is not that the participants were in some way deficient in higher order thinking (after all, the critical thinking contingent was, for some reason, the largest contingent at the November 1991 Workshop) or that the participants were in any way deficient in "advanced" communication skills. Rather, that particular construction of meaning for Goal 5 didn't show up because a context did not exist in which questions about the formulation of Objective 5 and about the evidential basis for the claim it signals could be asked: That context--the context of production for Objective 5--was simply not relevant to the context in which the text sign of Objective 5 was being used in November of 1991 and in which that text sign is still being used a year and more later. The context of production is simply not relevant to anything presently having to do with the NACSL. This latter point finds support in the facts that Emerson Elliott (who was then Acting Commissioner of NCES) could begin his "Foreword" (NCES, 1992, p. iii) to the NCES report on the 1991 Workshop by quoting the 1990 Objective 5 and that Sal Corrallo and Gayle Fischer could begin their "Background" (NCES, 1992, p. 3) section in the same volume with the above quotation from the National Goals Resource Group's March 1991, *Interim Report* and none of the three apparently be concerned about any disjunctions.

Neither, of course, at all relevant to the present context is the unflattering interpretation I hypothesized for Objective 5: The text sign that is Objective 5, having been accorded a political (and ideological?) stamp of approval from the Nation's governors and its President, has entered a very different context from the context of its production, a new context that the text sign itself helped shape once its social and political (and ideological?) status as a text sign was altered *vis a vis* the actions of the governors and the President, actions carried out verbally in the manner of a performative speech act (e.g., "I do thee wed") that, through its very utterance, altered not only the context but also the criterion of contextual relevance, which affected--in turn--what it is now possible to say or do effectively. The text sign of Objective 5 no longer exists in a context wherein and through which questions about why or about evidential bases for implicit claims can now be asked, but that text sign does exist in a context wherein debates about value and bases for implicit claims must yield in silence to the realities of a political action set in motion. Whether it's a good or bad action is simply not a question one can raise in the context of the NACSL development effort, which supports and values only those statements that can be construed or constructed as participating in and contributing to the action game already under way.

The writers of the November 1991 commissioned papers, as well as those persons who responded to the papers during the Workshop, appear to have been altogether sensitive to these very discourse-informing dimensions of context: To borrow one of the metaphors (cf. Greenwood, 1992) from the 1991 Workshop, the train has left the station; and the only questions now relevant are those having to do with which direction it will head, where it will ultimately stop, and how fast it will travel. Traveling now under the force of its own considerable (political) momentum, the train won't even have to stop for fuel.

The merits of the foregoing analysis of the situational and contextual dynamics of the NACSL development effort will, of course, have to be decided within the context of the stream of discourses that the present paper has entered, in the "context" of the "intertextual" world of which the present "text" is now a part (cf. Witte, 1992). It is a stream of discourses that has been flowing for sometime through a number of geographical and disciplinary regions, a stream that has gathered into itself not only Goal 5, Objective 5, but also a multitude of other texts (including Bush's 1988 campaign promise to become "the education President"), to which the NEGP goal can be seen as a response, and the multitude of other texts (for overviews, see Greenwood, 1992, and NCES, 1992) that the NEGP goal statements have themselves engendered, not all of which have been made an explicit part of the NACSL development effort.

To the extent that my analysis of the situation is accorded merit, that merit owes to the way of thinking and the way of speaking represented by the constructivist view of communication I have outlined. In fact, my "readings" of both the situation and the text sign of the Goal statements can only be regarded as constructivist readings, readings that depend--of course--on the very critical thinking skills that the current design of the NACSL development effort assumes are separable from "an advanced ability . . . to communicate effectively." As itself a text sign representing a construction of meaning predicated on prior texts and as itself a text sign that invites subsequent constructive readings, the present paper has become a part of an ongoing conversation, a conversation in which communication cannot be understood as the mere transmission of meaning, through the "container" of language, from a writer to his readers. That conversation, obviously constrained in some important ways, can be understood only as dialogue and interaction carried out through a complex network of statements and responses that are themselves executed through a multitude of texts--some "social" in the way I suggested previously, some oral, some exclusively printlinguistic (as I've deliberately made this one), and some a combination of alphabetic and nonalphabetic codes, etc.

It is a conversation during which participants (whether readers, writers, speakers, or listeners) constantly engage in meaning-constructive activities as those participants take their "turns" in the conversation (some "turns," such as the present text, are formally "ordered," while others are less so), in each case responding to previous discourses and in each case providing new occasions for subsequent responses by others. Moreover, the value of any utterance that enters the conversation cannot be determined independently of the conversation itself, independently of the stream of discourses of which it becomes a part. Whatever value a given statement may be said to have is situationally determined, and the value ascribed to a statement at one point in time may not be the same value ascribed to it at another point in time. Is text value, then, relative? Absolutely! And if the NACSL is really concerned with "*advanced* ability to . . . communicate effectively" (italics mine), value will not be represented as though it were not. In the world of "college graduates," just as in the world of the college graduates participating in the NACSL development process, value--like meaning itself--is, in very important and fundamental ways, determined constructively (i.e., dialogically and interactively) within particular contexts and situations. Like meaning, value is "relativized to scenes" or situations, and the only constant of a living scene is change itself.

From Theorizing and Practicing to Skills for Assessing

In concert, the more theoretical first part and the more (I hope) illustrative second part of the present section speak directly to both the cognitive and social complexity of communication in general and of the communicative acts of reading and writing in particular. By calling attention to that complexity, those two parts also speak, I think, indirectly to the sorts of understandings of communication and "text" that college graduates will need to develop and to the sorts of communication acts in which college graduates will need to engage if they are to "exercise the rights and responsibilities of citizenship" in a diverse society and world and to "compete" cross culturally in a "global economy": Apart from relationships of a personal nature, in no other participatory activities is it more necessary to understand, to be sensitive to, and to be able to act within the constructive dynamic of texts, contexts, and intertexts.

What, then, are the skills or, more properly, the categories of skills that "demonstrate" the college graduate's "advanced ability" to "communicate effectively" with regard to reading and writing?[3] Without intending to minimize the complexity or the importance of such "basic" skills as those associated with producing and comprehending words, sentences, and extended prose passages of different kinds, I would argue that such skills, while certainly embedded within any notion of "advanced ability" and while certainly necessary in order to "communicate effectively," are dimensions of reading or writing that need not and should not be a primary focus in an assessment of "advanced ability." My sense is that to do so not only sends out the wrong signals about "advanced ability" but also risks trivializing the assessment because of the multiple off-the-shelf tests and assessment methods currently available and used for assessing such "general" skills. One of the chief problems with such tests and methods, of course, is that they assume, among other things, that no important disjunctions obtain between, on the one hand, an artificial and tightly controlled assessment context (which includes the "test" instrument) wherein language is treated as an individual rather than a social phenomenon and, on the other hand, naturally occurring (and usually socially "messy") contexts of language production and use.[4] As the theoretical position I outlined and as the example I proffered suggest, neither such "general" skills nor the typical methods used in assessing them can distinguish between college graduates who can and cannot "communicate effectively" at an "advanced" level. In fact, it is the ability to deal effectively and appropriately with the very "social messiness" of language in naturally occurring situations and contexts that lies at the heart of my conceptualization of "advanced ability."

---

[3] More than the others, this third subsection differs from its counterpart in the original version of this essay discussed in Washington during the November 1992, Workshop. One change was organizational: In the present version, topics covered in the original third subsection are now distributed across the third and a fourth subsection. Several changes consist of elaborations on topics present in the original. To the extent that my revision of the subsection represents an improvement over the original, those improvements owe generally to discussions of the original paper among the Workshop participants. And they owe more specifically (a) to Marcia Farr's reminder (see her "Response to Daly and Witte Papers" attached as a Supplement to this document) that I should have addressed more explicitly such important issues as gender and ethnicity in my earlier treatments of process and situation and (b) to Bob Bracewell's (1992) and Mark Aulls's (1993) insightful and helpful critiques of what I had previously written about "skills" and their assessment. Apart from what I hope is a clearer and more detailed and considered representation of the "skills" I believe indicate "advanced ability," perhaps the most substantive difference between the present version and the earlier one is that, because of Bracewell's and Aulls's critiques, I now see as wrongheaded and indefensible my earlier call to assess the "skills" in an all-or-nothing manner.

[4] This is neither the time nor the place for a detailed exploration of other problems with traditional assessments, but I don't want to leave the impression that the one problem I've identified here is the only one. In fact, other problems stem from the artificiality of assessment contexts. Consider, for example, so-called "direct" assessments of writing. In "direct" assessments, people are typically required to produce a written text in response to a writing prompt that requires them to imagine themselves as writers with an imagined purpose in an imagined situation and writing for an imagined audience with imagined needs. One problem with such an approach is that it would appear, for example, to confound the ability to produce an appropriate and effective text with the ability to imagine a

To my way of thinking, "advanced ability" must refer to the ability to use reading and writing in the situationally and dynamically constructive ways I have suggested. Such use implies the need for college graduates to develop rather more sophisticated understandings of communication and "text" than I think many presently have--an understanding predicated on the multiple and variable interactions and intersections among context(s), text(s), and intertext(s), which contribute to the "social messiness" of specific situations and to the cognitive and social complexity involved in communicating appropriately and effectively within those situations. At the very least, that understanding would entail knowing ***what and why***

* the concept of "text" cannot be limited to linguistic "text,"

* acts of writing and reading "texts" involve both social and cognitive processes,

* producing and using "texts" are always in some sense collaborative acts,

* a "text" is always a response to at least one prior text, whether the prior text(s) be linguistic or nonlinguistic,

* a multitude of other "texts" may affect the way one reads or writes a text in a particular situation,

* the construction of "text" is also a construction of context and intertext,

* relations that obtain among context(s), text(s), and intertext(s) are reciprocal,

* a "text" is always about something,

---

situation within which one might do. In addition to raising this and other validity issues, such assessment practices also raise certain ethical issues not unlike those that are of concern among committees in academic institutions charged with deciding the ethics of using human subjects in research. In requiring people to respond to such prompts, "direct" assessments of writing not only require people to participate in a fabrication (which differs from a simulation in several important respects) but also to produce a fabrication. The ethical issue enters in when we ask ourselves whether requiring people to participate in such fabrications is not also teaching or reinforcing and thereby valorizing deception as ideology and rhetorical practice, that is, how to lie with words. In this regard, it is important to understand the difference between requiring people to produce such fabrications (inauthentic discourse?), on the one hand, and asking them what they would do "if" they were placed in the situation represented in the prompt or asking them to produce texts under the conditions of a true simulation.

- all symbol systems (including linguistic ones) can refer to something but can "contain" nothing,

- people can construct different meanings for a given "text,"

- writing and reading are not ends in themselves but are means to ends beyond themselves, and

- shared meaning is always constructed through "texts," that is, negotiated dialogically and interactively (i.e., socially) through the use of one or more symbol systems (for support of these dimensions, see Witte, 1992).

Given the foregoing thumbnail sketch of the bases for an adequate understanding of communication and "text," "advanced ability" can be defined as ***appropriately and effectively adjusting the processes, the topics, and the signs of one's communication to the demands of a given situation***. At first glance, this ability might appear to be an altogether commonplace and general one. And, in fact, it is, to some extent--particularly when we find ourselves in familiar situations with people we know. Indeed, we regularly adapt our communication to different situations: When we visit with our children at the dinner table, we talk in very different ways and about quite different topics than when we discuss with colleagues the latest article of interest in *The Educational Researcher*; similarly, when we write to loved ones, we typically write in a very different way than when we write a company memo about a pending change in health care coverage that will affect every employee. However, people do not always adjust their communication appropriately and effectively to the demands of specific situations, as the Clarence Thomas, Robert Packwood, and Marge Schott affairs indicate and as Marcia Farr illustrates in the Supplement attached to this essay. To be sure, many seemingly well-educated people remain ignorant of or insensitive to a host of, for example, gender and culture differences that influence the meanings constructed in particular situations and that, thereby, affect with increasing frequency judgments of the appropriateness and effectiveness of communicative acts. From my perspective, the ability to adjust one's communication to the demands of different situations is particularly crucial because, among other reasons, the U.S. and its institutions have come to look more like ethnic "salad bowls" than ethnic "melting pots" and the economies of humanity, environment, and commerce have become so obviously intertwined and interdependent globally. In addition, evidence such as that surveyed by Farr (see the attached Supplement) strongly suggests that adjusting one's communication to different situational demands is important because it can lead (a) to greater social harmony and (b) to savings in terms of the time and effort that is devoted to (mis)communication as communicants try to construct shared meaning(s). Yet it seems equally clear that this "ability"--which permits us to communicate appropriately and effectively through spoken and written texts in many familiar situations--does not always readily or even necessarily transfer to novel yet naturally occurring situations. An assessment aimed at "advanced ability" would need, it seems to me, to deal specifically with this issue.

As I have tried to suggest, the ability to adjust appropriately and effectively one's communication to situation-specific demands entails much more than being able to respond to, for example, "canned" writing prompts that might ask writers to address different imagined audiences, to write for different imagined purposes or in different assessor-determined forms or genres, or to treat certain assessor-determined contents or subject matters. As should be apparent, this "much more" points to a complex of skills, all of which I see as embedded in the overarching ability.

One such skill, or set of skills, can be referred to as ***"reading" (and interpreting) a given situation***. This skill involves recognizing conditions within or features of specific situations that are likely to affect the nature and success of communication. Such conditions are, of course, wide-ranging within and across specific occasions for both writing and speaking. Those conditions include variations in the gender- and culture-based "communication styles" Farr discusses (see attached Supplement), but they also include, for example, any number of possible status relations among communicants and any number of differences and similarities among communicants with regard to prior knowledge, perspectives on a given topic or issue, communication goals or purposes, and so forth. Such conditions or features, while representing potential threats to successful communication, need not preclude successful communication itself and may, in fact, constitute starting points for constructing new understandings or meanings, providing that honesty and goodwill among communicants prevail.

To be sure, the *ethos* of a speaker or writer and the ethical appeal of a text are never completely independent of other aspects of a specific situation and its context. (For an insightful discussion of *ethos* in discourse, see Cherry, 1988.) Yet it is important to understand that reading (and interpreting) situations in which goodwill and honesty among communicants can be assumed is a very different matter than reading (and interpreting) situations in which honesty and goodwill among participants cannot be assumed. Fortunately, many communicants are honest and bring an abundance of goodwill to different situations. But, unfortunately, not all. Accordingly, in some situations, perhaps in more than we feel comfortable thinking about, communication results not in the construction of shared meanings but of "mixed messages" or even outright deceptions that serve the ends of one communicant but not another. "Mixed messages" and outright deceptions are important in reading situations and hence in communication itself, and they suggest a skill that might be labeled ***reading signs of deception in a given situation***. Umberto Eco (1976) sees the study of signs (known as *semiotics*) as the study of "everything which can be used in order to lie" (p. 7). Because virtually everything can be used in some way to lie, Eco's definition suggests the importance and the complexity of reading signs of deception, a skill that even we who study semiotic theories and treatises and who think semiotically about communication often fail to apply to the communicative acts of particular moments. This is especially true when the goodwill and honesty we as individuals bring to those moments is accepted and used to another's advantage but never actually reciprocated in kind. However, just as situations may be imbued with signs of deception so, too, may they be imbued with signs of good will and honesty. And skill in ***reading signs of goodwill and honesty in a given situation*** is no less important than skill in reading signs of deception. The ethical dimension of communication that is the focus of these two skills is, of course, important not only with regard to the speakers and writers whose texts we hear and read but also with regard to our own representations of self when we speak and write. Accordingly, these two skills serve a "self"- monitoring function as well as an "other" monitoring function. Because deceptive communication practices are demonstrably often highly effective ones, what's at issue here, of course, is not effectiveness, but appropriateness.[5] My reasons for separating these two skills from skill in reading other dimensions of situation or context is that both require a sort of world knowledge that reading other dimensions of context do not and that both require making inferences about human character that seem to differ significantly from the inferences demanded by reading other dimensions of a situation. And I have separated these two skills from one another because the sorts of knowledge and inferences presupposed by each seem quite different: Perhaps there is some truth in the old adage that to be a good detective, you have to think like criminal.

The ways in which I have thus far represented components of "advanced ability" are potentially misleading because those representations could suggest to some an outside-in view of the relation between context or specific situation and writers (or speakers) and readers (or listeners). In this, those representations could engender the mistaken notion that with regard to writing (or speaking) "adjusting one's communication" entails, and only entails, accommodating presumably "writer-external" situational constraints through a sort of matching of text to context and intertext and doing so under conditions where text is seen as variable or malleable and where context or situation and intertext are seen as invariable, constant, or fixed. Those representations could also engender the mistaken notion that writers or speakers somehow exist apart from, are independent of, a context of readers or listeners in which discourse functions. But writers or speakers are part of context or situation as well: Context and situation cannot be defined as though it comprises some participants but not others. While it is certainly the case that elements of context or situation can influence or shape what writers (and speakers) do or say through texts as well as influence or shape what and how people read (or hear), it does not follow that context or situation can be understood simply in terms of what might exist outside the skin of, say, a writer or in terms of a one-way direction of influence from the outside in: Texts and their production influence and shape (i.e., construct) contexts as well, whether those contexts be associated with text production or use; and they invoke and contribute to intertextual streams of prior and subsequent texts.

---

[5] During an "authors' meeting" the evening before the November 1992, NACSL Workshop, I suggested that the NCES and its parent agency, the U.S. Department of Education, in developing an assessment that could alter educational practice substantially, might not want to make too much of the communication/critical thinking skills that would permit people to probe the ethical dimensions of communication because the U.S. political system (*vis a vis* election campaigns, for example) and the U.S. economic system (*vis a vis* advertising practices, for example) rely so heavily on deception for effective communication. A cynic might suggest that an educational system that would eliminate the use or the effectiveness of deceptive communication practices among politicians and advertisers would by itself lead to the complete collapse of the U.S. social system. A cynic might also predict that such a collapse is not imminent.

Indeed, if the present text is a successful one (i.e., appropriate and effective), it will at least contribute to the construction (or reconstruction) of the context within which the NACSL is being developed and to the construction of the context(s) within which it will be carried out. And if it succeeds in doing so, it will have done so in part because it gathers into itself various streams of discourse (intertexts), some of which are more accessible or salient to some readers than to others, and represents that combination in such a way that readers can anchor or ground what may be new information or new ideas to them in what they already know from the texts they have read and the "texts" they have experienced. At a more specific level, the writing of the original version of the present essay I regard as a constructive act through which I was able, in some respects, to (re)construct an audience (one which, I suspect, NCES personnel did not anticipate providing) for what I had to say about the nature of communication and what that implies about "advanced ability" and its assessment. And I also regard my writing of the original version as a constructive act through which I was able to construct, in some respects, other elements of context as well--a context that differed, as I suggested previously, from the one that was signaled by materials the NCES sent out in advance of my writing the essay. The constructed context was one that permitted those situated readers to construct shared ideas among themselves about communication and the assessment of "advanced ability," ideas that some readers not only thought to be novel but also ideas that some readers did not anticipate ultimately endorsing or holding themselves.[6] The importance of recognizing such constructive dimensions of communication lies in the fact that without them, new knowledge could not be created or shared among communicants. If contexts--and the audiences that in part define them--are nonameliorative, we would be limited to communicating only about what people already know in common. All of this is to say that writers not only ***author***ize ideas, they also ***author***ize the contexts in which their ideas are born and dimensions of the intertextual worlds in which those ideas can live. In short, ***author***ization of context and intertext is a lot of what writing (and speaking) is ultimately all about.

The context-constructive dimensions of communication clearly bear on one's ability to adjust his or her communication appropriately and effectively to the demands of a given situation. Part of what I did in writing the original version of this essay (and in writing the present revision) was to read (and interpret) situational demands as they were represented to me through documents from and conversations with NCES personnel and then to reconstruct those demands such that they opened a window of opportunity for me to represent a view of communication, "advanced ability," and the assessment of "advanced ability" that appeared not to figure in any important way in the conceptualization of the NACSL that was initially represented to me.[7] To the extent that I succeeded in the original version and to the extent that I succeed in the present revised version in identifying among those initial situational demands the conceptual and temporal space where my own ideas about communication, "advanced ability," and assessment of "advanced ability" can take shape and exist, my text(s) will have succeeded. ***Identifying and opening windows of constructive opportunity within a given situation***, or identifying and using "*kairotic* moments," as classical Greek rhetoricians would probably call them (cf. Kinneavy, 1986; Doheny-Farina, 1992; Miller, 1992), may be seen as another important dimension of "advanced ability" if the term *advanced* has anything to do with (re)constructing contexts in which new knowledge can itself be constructed and shared.

---

[6] The foregoing and following brief analyses regarding the reception of the original version of this essay find considerable support in what I gleaned from conversations with persons who read and then discussed in Washington the original version, from subsequent telephone conversations with some of those same people, and from face-to-face conversations, telephone conversations, and e-mail correspondence with a number of other persons who read the earlier version.

But the skills I have identified thus far point, I think, to only part of what "advanced ability" entails: One might well "read," interpret, and understand a situation "correctly" and recognize opportunities for constructing knowledge and context but, at the same time, have no skill in ***adapting or developing strategies for appropriate action predicated on one's knowledge of a given situation***. Moreover, this fifth component skill--or set of skills--seems to point to yet another: ***making informed and appropriate decisions in a given situation about what can or cannot be said, about what needs or needs not to be said, about how or how not to say it, and about when or when not to say it***.

Embedded within these "strategy" and "decision" skills is the recognition that college graduates are not a homogenous lot. College graduates, if they are anything, are a lot of different kinds of specialists: They are chemists, civil engineers, OERI employees, landscape architects, musicians, teachers, FBI agents, pilots, coaches, computer programmers, social workers, business managers, accountants, dieticians, farmers, insurance salesmen, policemen and women, etc. And their being specialists is necessary if they are to contribute to a "global economy." As specialists, college graduates must participate in specialized discourse communities, communities that are distinguishable both according to their specialized knowledges and according to their distinctive ways of thinking and speaking about the world. The "strategy" and "decision" skills thus point to being able to employ reading and writing in the meaning-constructive ways demanded by one's individual specialization within situations and contexts in which those specializations are acted upon and with. The strategy and decision skills are also intended to accommodate those occasions when college graduates are required to step outside their specialized discourse communities to communicate, in the context of carrying out work-related tasks, with persons in other specialized discourse communities.

---

[7] Evidence of the disjunction between my representation of communication--which most, if not all, the discussants of the original version of this essay shared or came to share--and the initial NCES representation is found, among other places, in the fact that the so-called "icon" that NCES created and distributed to discussants for representing and/or cataloguing communication skills was found by the discussants themselves to be woefully inadequate. I suspect that the discussants in Washington found the NCES "icon" (a grid-like affair) inadequate largely because of the difficulty of thinking in very sophisticated or theoretically sound ways about skills that might indicate "advanced ability" without also thinking about the situatedness of any communicative act that might entail "advanced ability."

In a culture and a political system whose engine often appears, more than anything else, to be an economic one, the "strategy" and "decision" skills also enable college graduates to perform in ways critical to citizenship. In fact, one reading of Goal 5 makes participation in the U.S. and local economies--neither of which are mentioned specifically--subordinate to participation in the "global economy" and citizenship. But citizenship should not be defined exclusively in economic terms (cf. Morris, 1989, 1992); and being a good citizen, which concept is itself--like communication--grounded in the social concept of community, demands stepping outside of one's workplace community or communities into a larger community whose individual members are much more heterogeneous and a larger community that may need expertise found in workplace settings but share neither the goals nor the methods of acting on goals that are found in workplaces. Accordingly, all of the skills I have identified can operate in the service of individual participation in a democratic society and its institutions, as well as within and across workplace settings.

Also viewed as a part of the "strategy" and "decision" skills is skill in using different symbol systems to construct meaning, whether the situation demands writing a text or reading a text. Although I deliberately made the present text a printlinguistic representation of my intended meaning, I could have "written" a number of pictorial representations of my ideas as well. In fact, some readers may prefer that I had done so. But I chose not to do so because every pictorial representation I was capable of constructing appeared to me to underrepresent woefully the complexity of the ideas I was trying to represent. My sense was, and is, that any representation that underrepresents that complexity at this stage in the development of the NACSL would increase the probability of an inadequate framework being constructed for the communication component. Yet the fact remains, many contemporary "texts" through which people communicate--from seating arrangements in meeting rooms, to labels on cereal boxes, to articles in scholarly journals, to telephone book Yellow Pages, to *Newsweek* articles, to Stephen Gould's *Wonderful Life* (1989)--all incorporate nonlinguistic symbol systems. The point is that such "texts" force reading and writing out of their historical and school-based moorings in printlinguistic symbol systems. Skill in constructing meaning through nonlinguistic symbol systems is part of what is presupposed by skill in making decisions about "how or how not to say" what one can or needs to say.

Throughout this essay, I have stressed the fact that neither language nor language users, whether they be readers or writers, exist in a social vacuum. Indeed, readers and writers are themselves important components of both situation and context. And the ameliorative character of context and intertext, together with the fact that every text is in some sense a response to a prior text, calls attention to the fundamental importance of ***collaborating with others during reading and writing in a given situation***, the seventh skill in my "preliminary listing." Even when we as individuals read, in the quiet of our studies or dens, texts such as those written for the 1991 NACSL Workshop or texts such as this one, we construct meaning collaboratively, in those cases by collaborating--through the text signs themselves--fairly directly with the authors of those texts. But in doing so, we also enter into collaborations with authors whose voices echo through the intertexts invoked as we construct meaning from the individual texts we read and try to anchor those meanings in what we already know. And the readers of our own texts go through a similar process, collaborating with us as writers through the texts we have written while invoking intertextual voices, which may or may not be exactly the same voices we invoked for ourselves as writers. But skill in collaborating with others extends well beyond interacting with them through the texts we read or write in private. Increasingly, collaboration during acts of reading or writing has become more dependent on face-to-face interactions as people in particular situations (e.g., the NACSL Workshops) collaborate through speaking and listening as they construct shared meanings from texts they have read and as they become constructive co-authors of texts (e.g., this revision) that are written. So important and pervasive is collaboration in its overt and covert forms (cf. Witte, 1992) that I would argue that the notion of a single author and how writing actually gets done are related in precisely the same way as the terms *sunrise* and *sunset* are related to a post-Copernican view of the universe: Both relational sets may give us some measure of anthropomorphic comfort but do little to alter existential reality. The more we learn about how writers work in naturally occurring situations, the more we understand how thoroughly dependent even texts such as this one, which lists but a single author, are on prior conversations with others regarding--at the very least--the ideas represented and authorized through the text itself.

The ability to adjust the processes, topics, and signs of one's communication to the demands of different situations implies, of course, knowing how to do so, which presupposes a certain level of metacognitive knowledge or skill. As I have suggested, some of this "how to" knowledge develops naturally through social interactions in familiar and repeatable situations and often without the aid of direct instruction, even though development can be and often is enhanced through feedback and facilitation peculiar to the situation at hand. Moreover, as Bracewell (1992) points out, the application of such knowledge in different situations may be either deliberate or nondeliberate, may entail either automatic or highly conscious decisionmaking. The earlier paragraphs in which I tried briefly to indicate some dimensions of the metacognitive knowledge I accessed in writing both the original and revised version of the present essay give, perhaps, some suggestion of the nature of that knowledge, as does the list of "that and why" knowledge which appears near the beginning of this subsection. And my brief analyses of my own processes in writing this and the earlier version of the present essay give, perhaps, some indication of one writer's skill, or lack thereof, in ***accessing and applying metacognitive knowledge of communication*** within what was for him an altogether novel writing situation. This eighth skill indicating "advanced ability" must, it seems to me, develop in parallel with the other seven. Moreover, it would seem that this skill can develop only to the extent that the others develop as well. At the same time, however, it is quite likely that it is the development of this eighth skill that permits development among the other seven.

In concert, the eight skills I have identified specify more completely the nature of "advanced ability" to "communicate effectively," which I defined as ***appropriately and effectively adjusting the processes, the topics, and the signs of one's communication to situational demands***. The eight skills--all components of "advanced ability"--are compatible with the "that and why" knowledge listed near the beginning of this subsection, a list which I argued represented the understandings of communication and "text" required to account for or explain the range of communicative acts in which college graduates must engage. Moreover, I indicated that the list of "that and why" knowledge was compatible with and derivable from both the constructivist model of communication I described and the extended illustration I provided of the model in action. Accordingly, if the theoretical argument and the grounding of the theory in practice are accepted, then it would seem to follow that my definition of "advanced ability" and my specification of the eight skills indicating that ability meet the twin criteria of theoretical compatibility and practical grounding. This latter point is, I think, an important one because it suggests how declarative knowledge of communication might be connected in practice with procedural knowledge. The relation of theory and practice is also important because any assessment of communication, whether of "advanced ability" or not, that is predicated on anything short of a necessary connection between grounded theoretical principles and specific skills will be difficult to justify and even more difficult to defend in any ethically satisfactory way.

Through Indirection Finding Directions Out: Skills, Performances, Levels

With regard to the design of any assessment involving human activity, it is appropriate to think about skills, performances, and levels of performances. Moreover, because assessments exist to distinguish among individuals or groups on the basis of performances in which specified skills are believed to have been enacted, it is important to consider the possible relations that obtain among the specified skills, the performances elicited, and levels of performance. And all of that is, of course, fairly straightforward--if you are dealing with easily defined and specified skills and if you are dealing with easily observed or measured performances or outcomes, such as driving a nail or punctuating sentences someone else wrote or "correcting" the grammar of a passage you read or solving a "story problem" in mathematics or writing a grammatically "correct" English sentence or ticking off from a list of possible choices the one indicating the main idea of a passage you read. But in an assessment whose focus is "advanced ability" to "communicate effectively," the possible relations that obtain among skills, performance, and levels for a given individual are a good deal more elusive and far more complicated.

In these regards, consider again the eight skills, or sets of skills, I identified and discussed in the previous subsection as indicators of "advanced ability":

1. ***Reading (and interpreting) a given situation***

2. ***Reading signs of deception in a given situation***

3. ***Reading signs of honesty and goodwill in a given situation***

4. ***Identifying and opening windows of constructive opportunity in a given situation***

5. ***Adapting or developing strategies for action predicated on one's knowledge of a given situation***

6. ***Making informed and appropriate decisions in a given situation about what can or cannot be said, about what needs or needs not to be said, about how or how not to say it, and about when or when not to say it***

7. ***Collaborating with others during reading and writing in a given situation***

8. ***Accessing and applying metacognitive knowledge of communication.***

In terms of assessment design--which entails fixing or setting relations among skills, performances, and levels--one of the difficulties, one of the complications, with the list of skills is that the individual skills, or sets of skills, do not appear to be discrete, to be independent of one another. That is to say, for example, that Skill 4 appears to be not unrelated to Skill 6, that Skill 2 probably has some bearing on performances associated with Skill 7, and that Skill 8 is likely directly related to the other seven. As I have already suggested, Skill 8 may be seen as both presupposing and being presupposed by the remaining seven skills.

Such observations point to another difficulty in translating the skills into assessment practice, namely, that it is impossible to organize the eight skills, or skill sets, as a conceptual hierarchy. Which skill, for example, would appear at the top of the hierarchy, and which at the bottom? Is ***reading signs of deception*** a necessary prerequisite for, say, effectively ***collaborating with others***? Perhaps. Under some circumstances. Is ***making informed and appropriate decisions*** a component of ***identifying and opening windows of opportunity***? Sometimes. But not always. Or is the latter dependent on the former? Probably. On occasion. Another difficulty is that--unlike driving a nail or punctuating somebody else's sentence--the skills cannot be ordered in a defensible "first this and then that and then that" sequence which, once invoked, would lead to successful communication. Any such sequencing of the entire list of skills could be easily deconstructed through appeals to handy and contradictory examples from actual practice. Additionally, the list cannot be ordered in such a way that it reflects a developmental sequence, whether like Shakespeare's seven ages of man or Piaget's developmental stages. And these difficulties are exacerbated by the fact that there is, as far as I know, no body of research that would permit us to deal with them in a very sophisticated way.

Now if, as I have argued, "advanced ability" comprises the complete set of eight skills or skill categories, it is tempting to argue that "advanced ability" can be signaled only in those cases where all the stated and implied skills contribute equally and simultaneously to success in communication. With some *a priori* specification of success as the principal performance criterion, one might even be inclined to argue that "all or nothing" judgments of performance with regard to the skills would be in order. That is to say, it might make some intuitive sense to think of "advanced ability" as comprising all eight skills or categories of skill, with their employment being at a maximum level in the service of a single communicative act. However, a number of problems speak against that view and, *mutatis mutandis*, against organizing an assessment of "advanced ability" around such a view.

One of the problems is that of finding a situation, an occasion, for writing and reading (or for either writing or reading) wherein the demands for all eight skills at some maximum level would be equal. Finding such a situation would be tantamount to discovering for communication assessment a philosopher's stone that would turn base metal in gold. It is highly unlikely that any naturally occurring situation, or any contrived one, could elicit a performance that demanded the use of all eight skills at a maximum level. Indeed, one of the principal arguments of this essay is that different situations make different demands on individual performance. That is to say, communication is always situated in a social context. For example, "windows of constructive opportunity" are not available in every situation, nor is it always possible to construct a "window" when one is not available. Accordingly, skill in "identifying and opening" such windows will not figure importantly in every situation. But that fact does not make ***identifying and opening constructive windows of opportunity*** any less an indicator of "advanced ability." Similarly, ***reading signs of deception*** is not a skill that one would necessarily need or use in every situation, however high the costs might be of not using the skill at some maximum level during those moments that truly require it. Likewise, not every situation demands or warrants ***collaborating*** with others overtly ***during reading or writing***. A second problem stems from the first: If a situation cannot be imagined or constructed wherein demands for maximum performance were equal across skills or skill categories, then it is difficult to see how "advanced ability" could even be scored in an "all-or-nothing" manner, to say nothing about the difficulty of interpreting and generalizing from such scores. A third problem that attends an "all-or-nothing" judgment is that (a) very little information could be derived from it that would enable the examinee to improve his or her performance and (b) such a judgment could contribute little or nothing to the design and development of instructional practice that could prepare students for the assessment or that could provide remediation for those judged "unsuccessful." In short, it is unlikely that the skills, or sets of skills, could be properly judged in an "all-or-nothing" manner.

If an "all-or-nothing" judgment of performance is neither sound nor possible, then it might appear that rendering judgments of performance with regard to each of the eight skills or sets of skills would provide a suitable alternative. But such an approach is not itself devoid of problems, both with regard to defining and setting performance levels and with regard to determining what--according to those specified levels--separate judgments for each skill might reveal about the "advanced ability" evidenced by an individual performance in a particular situation.

Defining and setting performance levels at the present time would appear to be a very risky and probably foolish undertaking, largely because it would have to be accomplished without an adequate grounding. As the March 1991, *Interim Report* of the National Education Goals Resource Group indicates, "neither national nor state information is currently available on the ability of college graduates to . . . 'communicate effectively'" (quoted from NCES, 1992, p. 3). Even though some studies of, particularly, the writing of college graduates have been conducted and point to, as I have indicated, the sorts of skills I have specified, results of those studies are largely exploratory in nature and cannot be seen as contradicting the conclusion of the *Interim Report*. Moreover, none of the studies addresses the question of performance levels specifically. Nor is much help to found in documents such as the Department of Labor's *Learning a Living* (SCANS, 1992), which deals not with "advanced ability" as I have represented it but, rather, with "survival" or "basic" performances expected of individuals entering different sorts of jobs. If the skills or sets of skills I have identified have theoretical and practical merit, then what is demanded in advance of defining and setting performance levels for the skills is a programme of research that systematically examines, in workplace settings, "advanced ability" as I have defined it in terms of the eight skills or sets of skills. Such a programme would most properly study a sample of recent college graduates stratified according to Department of Labor information regarding the number of college graduates employed in different jobs within the various job categories. Relying primarily on observational techniques and data, the research programme would document performances associated with the eight skills or skill categories and derive from those data both performance criteria and performance levels that might be applicable to the assessment itself. In my judgment, it is only through such an effort that the necessary performance levels could be defined and set such that they would be credible as well as useful in assessment practice.

Defining and setting performance levels that are both credible and useful addresses, however, only one of the two problems associated with assessing the skills individually. The second problem is a different sort of practical problem than not having at hand a body of reliable information. If, as I have argued and (I hope) demonstrated, communicative acts reflecting "advanced ability" are always tied to particular situations and if "advanced ability" itself is situational rather than strictly dispositional, then it would follow that judgments of levels of performance with regard to each skill or skill set according to specified performance criteria would also be situational in nature. That is to say, just as individual performance is a function of situational demands on performance, so also would judgments of, say, "success" and "nonsuccess" be of function of the situation in which the judged performance occurred. There is no reason to believe that human judgment is any less "relativized to scenes" than communication performance or meaning itself. In addition to the question raised by the fact that a given situation for communication would not likely elicit performances that could be judged in terms of all skills or skill categories, the fact of judgments "relativized to scenes" raises the practical question of how many situated performances of a given individual must be judged before that individual can be said to have demonstrated "advanced ability."

I titled the present essay "No Guru, No Method, No Teacher" largely because the history of large-scale assessments provides no certain answers for the conundrums raised in the present section. My sense is that to the extent that a communication component will figure importantly in the NACSL, the NACSL will need to deal systematically and carefully with the questions and issues I have raised about performance levels. The answers and solutions will not be easily had, nor will finding them be an inexpensive undertaking. But finding the answers and solutions the NACSL must. What's at stake is nothing less than credibility and validity.

Notes Toward a Communication Assessment Model

Whereas the four parts of the previous major section respond to two of the directives to authors, the present section responds to the third: "to address and make recommendations in regard to . . . a format or approach to assessment of the skills in the skill area" (Corrallo, 1992, p. 5). The previous subsections are, as I stated early on, predicated on the assumption that what is and can be assessed by way of skills indicating "advanced ability to . . . communicate effectively" is affected by how assessment is carried out. Although I've stated that assumption as an assumption, I regard its substance as axiomatic. As I hope I've been able to make clear, testing for narrow, isolated skills associated with "communicating effectively" cannot measure "advanced ability" in any theoretically or practically defensible way. Any assessment indicating "advanced ability" with regard to the communication component of the NACSL will have to capture the complex dynamics I have tried to suggest in the previous major section by stressing implicitly throughout the theoretical and practical significance of the etymological relation of the term *communication* to the term *community*.

What this etymological, theoretical, and practical connection implies is that any performance elicited in the service of communication assessment will have to be elicited in such a way and evaluated in such a way as to respect the complex dynamic of context, text, and intertext in which authentic readers and authentic writers participate as a matter of course. Hence the sort of communication assessment I envision for the NACSL is a performance assessment that meets rather stringent requirements for authenticity--authenticity with regard to tasks, texts, settings, communicators, and evaluators. Authenticity with regard to any of these components exists, of course, as a continuum, from the totally inauthentic or contrived (as in the manner of multiple-choice tests of reading and writing skills and so-called "direct" assessments of writing) to the completely authentic, which--in the case of college graduates--would have to be found "on the job." The "completely authentic" is probably not practical, but the "totally inauthentic and contrived" can be avoided.

A medial approach could, I think, be something like the "curriculum event" approach to communication arts performance assessment that I am trying to develop, with the considerable help of teachers and colleagues, for the Wisconsin Student Assessment System (Witte & Vander Ark, 1992). An "event" approach to communication assessment makes some sense because it incorporates and capitalizes on the necessary relationship between communication and community, a relationship that is either ignored or trivialized in other forms of communication assessment. Such an "event" can be defined as a series of theoretically and practically coherent activities structured or organized such that they lead to performances of the sort suggested by the eight "selected" skills or sets of skills I have included in my "preliminary listing." As I think about such an event, it could be seen as a sort of capstone, interdisciplinary "course" or "class" that students take during their final semester of college.

The students, who would come from different disciplinary orientations or specializations and who would represent demographically the local population of students, enrolled in the "class" would be asked to identify and agree upon, through negotiations with one another, a significant problem to solve, a problem that would admit exploration *vis a vis* activities requiring the use of reading and writing from a variety of disciplinary perspectives. One of the requirements would be that the group-generated solution be presented to some group, agency, or person--probably outside the academic institution itself--capable of acting on (i.e., implementing) the group solution. Assessment data--both of a descriptive and an evaluative sort--would be collected throughout the entire process and would focus on both individual and group performances.

The notion of an assessment as an "event" that unfolds over an obviously protracted period of time and how it would differ from more traditional forms of assessment might be better understood through an extended and compound analogy, one involving some aspects of the history of popular music and one giving additional point to my borrowing part of the title for the present paper from Van Morrison's "No Guru, No Method, No Teacher." Before rock and roll began to influence substantially the popular music scene during the mid-1950s, popular radio had committed itself to playing recordings whose lengths corresponded roughly to the attention span of a large segment of its listening public and whose lengths paralleled roughly the stretch of available time that audiences generally had for active listening. And with regard to popular radio, the situation is not a lot different now. Radio play, as the principal avenue to financial success for musicians and recording companies, has essentially defined most popular music as a succession of three-minute wonders--short, highly focused, and (usually) danceable songs. And even when new forms of popular music emerge, those forms are typically forced into the straightjacket of the "three-minute-wonder."

The emergence and acceptance of rock and roll on the popular-music scene during the mid-1950s provides a case in point. Because of its repeated radio play and its consequent overwhelming popularity and commercial success, one such three-minute wonder (Elvis Presley's 1956 cover of "Heartbreak Hotel" [e.g., Presley, 1958, 1973]) effectively defined the temporal limits of the genre of the "pop" rock-and-roll song for nearly a decade. In short, commercial rock and roll during the mid-1950s and after was, like its Tin Pan Alley predecessors (e.g., Perry Como's 1952 "Don't Let the Stars Get in Your Eyes" [Como, c. 1959]), defined by time limits considered acceptable for commercial radio.

Because commercial recording success and popular taste were and are directly related to radio play, the three-minute wonder has dominated popular music generally and rock and roll specifically in spite of the availability from about 1950 onward of a medium--the long-play album--that provided the technological potential for alternative conceptualizations of popular music. However, when popular recording artists (whether the Presleys or the Comos) of the mid-1950s began recording on long-play albums, the "new" medium had virtually no effect on either the nature or the duration of the popular songs released. Long-play albums of the period never amounted to anything other than anthologies of largely unrelated three-minute wonders. In fact, the creative and artistic potential afforded by the "new" medium was not tapped in the popular music field until singer-songwriter Bob Dylan released "The Times They Are a-Changin'" in 1964, a folk-revival album that now seems quite dated but that, nevertheless, marked the first time a popular recording artist took creative advantage of the "new" recording medium and released a collection of songs that were thematically and musically coherent and that departed from the norm of the three-minute wonder by offering songs up to seven-plus minutes in length ("With God on Our Side").

To be sure, the three-minute wonder has dominated popular music (there have, of course, been a few noteworthy exceptions such as The Beatles' enormously popular "Hey, Jude" [e.g., The Beatles, 1970, 1973], which runs over seven minutes) because of the role radio plays in defining popular taste; but a number of popular artists and groups--principally rock-and-roll artists (e.g., Graham Parker, Van Morrison, The Beatles, Joni Mitchell, Jackson Browne, Stevie Wonder, Pink Floyd, Carole King, Jethro Tull, Paul Simon, Bruce Springsteen, Neil Young, The Who)--have followed Dylan's lead and example (which he himself has repeated on subsequent albums) in making creative and artistic use of the long-play album and its modern successors--the audio cassette and the CD--by recording albums that stand up well as integrated wholes. Such albums urge a conceptualization of rock and roll that departs from the status quo definition of the single song as a three-minute wonder and of the album as a collection of unrelated or marginally related singles.

The point of the foregoing is that, like contemporary rock and roll, performance assessment--which is hardly a new idea (recall Horace Mann's arguments during the 1860s for the use of written examinations in the Massachusetts public schools)--requires a new definition. Historically, but particularly since commercial testing became something of a growth industry, performance assessment "tasks" and assessments generally have been defined in much the same way as the three-minute wonder defines most conceptualizations of rock and roll music, and performance assessment "instruments" have been defined as collections of "three-minute wonders," the notorious "items" and "item sets" that test makers invoke in the process of constructing a distinctive "test world." The difficulty is that communication can neither be described nor molded to fit nicely or comfortably into a "test world," into either a three-minute wonder or an anthology of three-minute wonders. Communication, as I've tried to indicate throughout this paper, is the world of the integrated long-play album, the world of "album-oriented rock." Yet current communication assessment practice assumes something like the three-minute wonder with regard to the processes of language production and use.

From my perspective, traditional communication assessment practice--whether it valorizes three-minute wonders or collections of three-minute wonders--creates validity-threatening disjunctions among the constructed "test world," authentic language performances, and best educational practices--the latter two having, in my judgment, more in common with the integrated long-play album than with either the three-minute wonder or collections of them. The capstone, interdisciplinary course that I envision is one that defines performance assessment not in terms of a three-minute wonder or a string of three-minute wonders, but in terms of an event whose nature is more holistic than meristic, more integrated than fragmented.

The explanatory power of any analogy is, of course, limited: Analogies always break down at some point because the points of correspondence are never exact. When they do break down, explanations must advert to the concrete particulars of the object or phenomenon under discussion. The case of assessment as a protracted event is no different; and much more by way of conceptualization is in order. Indeed, most of the details of such an assessment approach would need to be worked out. But the advantages of the approach are that it would engage advanced college students in, first, the sorts of authentic languaging activities that are embedded within the eight skills or categories of skills I have listed, that it would engage students in authentic tasks with authentic others, and that it would be totally defensible from an educational perspective. Such a protracted assessment event would, if I can reach anaphorically across the foregoing, truly represent a horse and not a camel, nor a *hamel* nor a *corse*.

Some After Words

Given Sal Corrallo's tripartite set of directives, the present paper is intended as a small contribution to an ongoing conversation--a conversation in which I, like many of my readers, became a participant well after the conversation began. And in writing the present paper, I have been constantly reminded of how I felt as a boy between 3 and 4 years old when I, startled by the sounds issuing forth from grown-ups in an adjoining room, would suddenly awaken from an afternoon nap. And I remember on occasion standing on the mattress of the bed with my arms and elbows propped on the side rails and listening intently to the sounds, often muffled and indecipherable, I'd hear coming from the next room--trying to make some sense of words I did recognize, trying to infer presences from known voices, trying to determine intentions from inferable presences, all while I was, I assume, trying to decide whether the image of the conversational context I was constructing mentally was one that I wanted to enter. There were, it must be understood, certain risks associated then with entering conversations at inappropriate times, particularly when your father was a minister and the living room in your home functioned as an undeclared refuge for distraught parishioners who had--I would later learn--just received word that a loved one had been wounded, declared missing in action, or killed in some distant place whose name for a small boy did little more than mark a severe distance between sound and sense. During those times, experiences--about which I'm fully incapable of wrapping words--taught me a lot about the risks that attend one's walking in on ongoing conversations, not the least of which is saying the wrong thing at the wrong time.

Whether I've used the conversational turn represented by the present paper to say all the wrong things or to say some right things in the wrong context are matters that will ultimately be determined by the subsequent turns taken during the NACSL conversation. My hope is that my text has helped construct a context that will permit at least some parts of that text to be heard among the many able voices that have become, and among those that will subsequently become, parts of an enormous and rich intertext that is emerging from the ongoing NACSL conversation.

References

Alexander, P.A., Schallert, D.L., & Hare, V. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research, 61*, 315-343.

Aulls, M.W. (8 January 1993). Telephone conversation with author.

Bataille, R.R. (1982). Writing in the world of work: What our graduates report. *College Composition and Communication, 33*, 226-283.

Barnum, C.M., & Fischer, R. (1984). Engineering technologists as writers: Results of a survey. *Technical Communication, 31* (2), 9-11.

Bazerman, C. (1983). Scientific writing as social act: A review of the literature on the sociology of science. In P.V. Anderson, J. Brockman, & C. Miller (Eds.), *New essays in technical communication: Research, theory, and practice*. Farmingdale, NY: Baywood.

Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication, 2*, 3-23.

The Beatles. (1970). *Hey Jude: The Beatles again*. Catalogue No. SW-385. New York: Apple Records.

The Beatles. (1973). *The Beatles/1967-1970*. Catalogue No. SBEX-11843. Hollywood, CA: Capitol Records.

Beers, T. (1987). Schema-theoretic models of reading: Humanizing the machine. *Reading Research Quarterly, 12*, 369-377.

Berkenkotter, C. (1991). Paradigm debates, turf wars, and the conduct of sociocognitive inquiry in composition. *College Composition and Communication, 42*, 151-169.

Berkenkotter, C., Huckin, T., & Ackerman, J. (1987). Conventions, conversations, and the writer: Case study of a student in a rhetoric Ph.D. program. *Research in the Teaching of English, 22*, 9-44.

Berkenkotter, C., Huckin, T., & Ackerman, J. (1991). Social contexts and socially constructed texts. In C. Bazerman & J. Paradis (Eds.), *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities*. Madison: University of Wisconsin Press.

Bickard, M. (1980). On models of knowledge and communication. In M. Hickmann (Ed.), *Proceedings of a working conference on the social foundations of language and thought*. Chicago, IL: University of Chicago, Center for Psychosocial Studies.

Bizzell, P. (1982a). College composition: Initiation into the academic discourse community. *Curriculum Inquiry, 12*, 191-207

Bizzell, P. (1982b). Cognition, convention, and certainty: What we need to know about writing. *Pre/Text, 3*, 213-243.

Bracewell, R. J. (10 December 1992). Personal correspondence with author.

Bracewell, R.J., & Breuleux, A. (*in press*). Substance and romance in the analysis of think-aloud protocols. In P. Smagorinsky (Ed.), *Verbal reports in the study of writing: Problems and potentials*. Newbury Park, CA & London: Sage.

Bracewell, R.J., Fredriksen, C.H., & Fredriksen, J.D. (1982). Cognitive processes in composing and comprehending discourse. *Educational Psychologist, 17*, 146-164.

Brandt, D. (1986). Text and context: How writers come to mean. In B. Couture (Ed.), *Functional approaches to writing: Research perspectives*. Norwood, NJ: Ablex.

Brandt, D. (1992). The cognitive as the social: An ethnomethodological approach to writing process research. *Written Communication, 9*, 351-355.

Breland, H.M., & Gaynor, J.L. (1979). A comparison of direct and indirect assessments of writing skill. *Journal of Educational Measurement, 16*, 119-128.

Brodkey, L. (1988). *Academic writing as social practice*. Philadelphia, PA: Temple University Press.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical review. *Research in the Teaching of English, 18*, 65-81.

Cherry, R.D. (1988). *Ethos* versus *persona*: Self-representation in written discourse. *Written Communication, 5*, 251-276.

Como, P. (c.1959). *Como's golden records*. Catalogue No. LPM-1981. Montreal PQ: RCA Records.

Corrallo, S. (1992). Background paper for meeting participants, authors, and reviewers. (Revised August 28, 1992). Washington, D.C.: National Center for Education Statistics.

Diederich, P.B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Doheny-Farina, S. (1992). The individual, the organization, and *kairos*: Making transitions from college to careers. In S.P. Witte, N. Nakadate & R.D. Cherry (Eds.), *A rhetoric of doing: Essays in honor of James L. Kinneavy* (pp. 293-309). Carbondale: Southern Illinois University Press.

Dunbar, S.B. (1991). On the development of a national assessment of college student learning: Measurement policy and practice in perspective. Paper commissioned by NCES for the Higher Order Thinking and Communication Skills Study Design Workshop. Washington, D.C.

Dylan, B. (1964). *The times they are a-changin'*. Catalogue No. CS 8905. New York: Columbia Records.

Eco, U. (1976). *A theory of semiotics*. Bloomington: Indiana University Press.

Faigley, L. (1985). Nonacademic writing: The social perspective. In L. Odell & D. Goswami (Eds.), *Writing in nonacademic settings*. New York: Guilford.

Faigley, L. (1986). Competing theories of process. *College English*, *48*, 527-542.

Faigley, L., Miller, T.P., Meyer, P.R., & Witte, S.P. (1981). *Writing after college: A stratified survey of the writing of college-trained people*. Writing Program Assessment Project Technical Report No. 1 (FIPSE Grant G008005896). Austin: University of Texas. ERIC Document 210 708.

Farrell, T. (1983). IQ and standard English. *College Composition and Communication*, *34*, 470-484.

Fillmore, C.J. (1977). Topics in lexical semantics. In R.W. Cole (Ed.), *Current issues in linguistic theory* (pp. 76-138). Bloomington: Indiana University Press.

Flower, L. (1989). The construction of purpose in reading and writing. *College English*, *50*, 528-550.

Flower, L. (in press). *The construction of negotiated meaning: A social cognitive theory*. Carbondale: Southern Illinois University Press.

Flower, L., & Hayes, J.R. (1984). Images, plans, and prose: The representation of meaning in writing. *Written Communication*, *1*, 120-160.

Flower, L., Schriver, K., Carey, L., Haas, C., & Hayes, J.R. (1992). Planning in writing: The cognition of a constructive process. In S.P. Witte, N. Nakadate, & R.D. Cherry (Eds.), *A rhetoric of doing: Essays on written discourse in honor of James L. Kinneavy*. Carbondale: Southern Illinois University Press.

Flower, L., Stein, V., Ackerman, J., Kantz, M.J., McCormick, K., & Peck, W.C. (1990). *Reading to write: Exploring a cognitive and social process*. New York: Oxford University Press.

Freedman, S.W., & Calfee, R.C. (1983). Holistic assessment of writing: Experimental design and cognitive theory. In P. Mosenthal, L. Tamor, & S.A. Walmsley (Eds.), *Research on writing: Principles and methods*. New York: Longman.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Geisler, C. (1991). Toward a sociocognitive model of literacy: Constructing mental models in a philosophical conversation. In C. Bazerman & J. Paradis (Eds.), *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities*. Madison: University of Wisconsin Press.

Godshalk, F.I., Swineford, F., & Coffman, W.E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.

Gould, S.J. (1989). *Wonderful life: The Burgess Shale and the nature of history*. New York: Norton.

Greene, S. (1990). Toward a dialectical theory of composing. *Rhetoric Review, 9*, 149-172.

Greene, S. (1992). Mining texts in reading to write. *Journal of Advanced Composition, 12*, 151-170.

Greenwood, A. (1992). *National assessment of college student learning: Getting started*. Unpublished paper from the National Center for Education Statistics.

Kinneavy, J.L. (1986). *Kairos*: A neglected concept in classical rhetoric. In J.D. Moss (Ed.), *Rhetoric and praxis: The contribution of classical rhetoric to practical reasoning* (pp. 79-105). Washington, D.C.: Catholic University Press of America.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lenth, C.S. (1991). The context and policy requisites of national postsecondary assessment. Paper prepared for the Office of Educational Research and Improvement, U.S. Department of Education. Washington, D.C.

Malinowski, B. (1946). The problem of meaning in primitive languages. In C.K. Ogden & I.A. Richards, *The meaning of meaning: A study of influence of language upon thought and of the science of symbolism* (Supplement 1, pp. 296-336). New York: Harcourt, Brace & World. (Original work published 1923)

Mann, H. (1845). Boston grammar and writing schools. *The Common School Journal*, 7 (21-23), 321-368.

Miller, C.R. (1992). *Kairos* in the rhetoric of science. In S.P. Witte, N. Nakadate & R.D. Cherry (Eds.), *A rhetoric of doing: Essays in honor of James L. Kinneavy* (pp. 310-327). Carbondale: Southern Illinois University Press.

Morris, S.W. (1989). *Renewing civic capacity*. ASHE-ERIC Higher Education Report, No. 8. (Photocopy supplied to author by NCES without additional publication information.)

Morris, S.W. (1992). Skills for citizenship. Paper prepared for the U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.

Morrison, V. (1986). *No guru, no method, no teacher.* Catalogue No. 830-077-1. St. Laurent, Quebec: Mercury/Polygram Records.

Myers, G. (1985). The social construction of two biologists' proposals. *Written Communication*, 2, 219-245.

Myers, G. (1990). *Writing biology: Texts in the social construction of knowledge*. Madison: University of Wisconsin Press.

National Center for Education Statistics. (1992). *National assessment of college student learning: Issues and concerns*. NCES 92-068. Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

National Education Goals Panel. (1991). *America 2000: An education strategy*. Rev. Ed. Washington, D.C.: U.S. Department of Education.

Nystrand, M. (1986). *The structure of written communication: Studies in reciprocity between writers and readers*. Orlando, FL: Academic Press.

Nystrand, M. (1989). A social-interactive model of writing. *Written Communication, 6*, 66-85.

Nystrand, M., Greene, S., & Weimelt, J. (in press). Where did composition studies come from? *Written Communication.*

Oxenham, J. (1980). *Literacy: Writing, reading, and social organization.* London: Routledge & Kegan Paul.

Presley, E. (1958). *Elvis' golden records.* Catalogue No. LPM-1707. New York: RCA Records.

Presley, E. (1973). *Elvis: A legendary performer*, vol. 1. Catalogue No. CPL1-0341. New York: Race Records.

Rabinowitz, P.J. (1989). Whirl without end: Audience-oriented criticism. In G.D. Atkins & L. Morrow (Eds.), *Contemporary literary theory.* Amherst: University of Massachusetts Press.

SCANS (1992). *Learning a living: A blueprint for high performance.* Washington, D.C.: U.S. Department of Labor, the Secretary's Commission on Achieving Necessary Skills.

Schallert, D.L. (1982). The significance of knowledge: A synthesis of research related to schema theory. In W. Otto & S. White (Eds.), *Reading expository material.* New York: Academic Press.

Sebeok, T.A. (1991). *A sign is just a sign.* Bloomington: Indiana University Press.

Spiro, R. (1980). Constructive processes in prose comprehension and recall. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, NJ: Lawrence Erlbaum.

Spivey, N.N. (1984). *Discourse synthesis: Constructing texts in reading and writing.* Newark, DE: International Reading Association.

Spivey, N.N. (1987). Construing constructivism: Reading research in the United States. *Poetics, 16*, 169-192.

Spivey, N.N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*, 256-287.

Stine, D., & Skarzenski, D. (1979). Priorities for the business communications classroom: A survey of business and academe. *Journal of Business Communication, 16* (3), 15-30.

Storms, C.G. (1983). What business school graduates say about the writing they do at work: Implications for the business communication course. *ABCA Bulletin, 46* (4), 13-18.

Stubbs, M. (1980). *Language and literacy: The sociolinguistics of reading and writing*. London: Routledge & Kegan Paul.

Tierney, R.J., & Pearson, P.D. (1983). Toward a composing model of reading. *Language Arts, 60*, 568-580.

Tierney, R.J., LaZansky, J., Raphael, T., & Chen, P. (1987). Authors' intentions and readers' interpretations. In R.J. Tierney, P. Anders, & J. Mitchell (Eds.), *Understanding readers' understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Venezky, R. (1991). Assessing Higher Order Thinking and Communication Skills: Literacy. Commissioned paper prepared for the NCES Workshop on the Assessment of Higher Order Thinking and Communication Skills of College Graduates.

Vygotsky, L. (1934/1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT Press.

White, E.M. (1985). *Teaching and assessing writing*. San Francisco: Jossey-Bass.

White, E.M. (1991). Assessing Higher Order Thinking and Communication Skills in College Graduates Through Writing. Commissioned Paper prepared for the NCES Workshop on the Assessment of Higher Order Thinking and Communication Skills of College Graduates.

Willingham, J.R. (1989). The new criticism: Then and now. In G.D. Atkins & L. Morrow (Eds.), *Contemporary literary theory*. Amherst: University of Massachusetts Press.

Witte, S.P. (1991a). Writing in context: A prolegomenon to a constructivist semiotic. Paper presented during the annual convention of the American Educational Research Association. Chicago.

Witte, S.P. (1991b). Situated writing: An exploration. Invited lecture, Benton Center Seminar Series, Department of Education, University of Chicago.

Witte, S.P. (1991c). Writing in context: Writing ability within a constructivist semiotic framework. Invited Lecture, Faculty of Education Seminar Series and the Explorations in Cognitive and Neuro Sciences Colloquium Series, McGill University. Montreal, P.Q.

Witte, S.P. (1991d). Context, text, and intertext: Writing in a semiotic framework. Invited Lecture, Department of Linguistics Colloquium Series, Carleton University. Ottawa, Ontario.

Witte, S.P. (1992). Context, text, intertext: Toward a constructivist semiotic of writing. *Written Communication, 9*, 237-308.

Witte, S.P., Cherry, R.D., Meyer, P.R., & Trachsel, M. (*in preparation*). *Holistic evaluation of writing: Issues in theory and practice.*

Witte, S.P., Nakadate, N., & Cherry, R.D. (1992). Introduction. In S.P. Witte, N. Nakadate, & R.D. Cherry (Eds.), *A rhetoric of doing: Essays on written discourse in honor of James L. Kinneavy*. Carbondale: Southern Illinois University Press.

Witte, S.P., Trachsel, M., & Walters, K. (1986). Literacy and the direct assessment of writing: A diachronic perspective. In K.L. Greenberg, H.S. Wiener & R.A. Donovan (Eds.), *Writing assessment: Issues and strategies*. New York: Longman.

Witte, S.P., & Vander Ark, C. (1992). WSAS communications/language arts performance assessment: The Wausau workshop. Working paper. Madison: Wisconsin Center for Education Research, University of Wisconsin.

SUPPLEMENT

(I am grateful to Professor Farr for giving me permission to print her response, to the original version of my essay, as a ***SUPPLEMENT*** *to the revised version of my paper. The time allotted for the preparation of my original paper did not permit me to treat directly the important issues of gender and ethnicity that Professor Farr addresses specifically in her response. Professor Farr's response was prepared for the U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics' Workshop on Assessment of Higher Order Thinking and Communication Skills of College Graduates, November 17-18, 1992.--S.P.W.)*

Response to Daly and Witte Papers

Marcia Farr
Dept. of English, m/c 162
University of Illinois at Chicago
Box 4348, Chicago, IL 60680

Although these two papers synthesize a great deal of research and theory that bears on the assessment of important communication skills, neither one treats a central issue in the Department of Education's concern about the communicative abilities of the current and future workforce of this country. Underlying the national concern about the educational preparedness of the workforce is the recognition of a substantial demographic shift taking place in the population both of the country as a whole and of new workers entering the workplace. In the country as a whole, some estimates predict that within 100 years, non-Hispanic European-origin people will constitute a minority in the United States (Sima, 1987). And as we approach the 21st century, new workers are increasingly comprised of women and minorities. A major part of the concern about communication abilities, then, is based on the fact that both culture- and gender-based differences in communicative styles are conflicting with traditionally male and European ways of communicating both in the workplace and in educational institutions.

While I argue that such differences in communicative style must be taken into account in assessment procedures, I do not argue for total relativity toward communicative differences. Rather, decisions about what kinds of communicative styles are valued in the workplace and for democratic participation in the society can be made on the basis of effectiveness in particular contexts--and effectiveness presumes inclusion of all workers and students, not exclusion, since it is clear that everyone is needed to comprise a fully skilled and educated society. For example, both papers mentioned the importance of collaborative processes in the workplace (and, indeed, businesses that are termed "high performance" workplaces rely on such collaboration as part of shared responsibility and authority). Such collaboration, however, is a recent innovation, and, in fact, may be congruent with the communicative styles of females and several minority groups. Ideally, then, decisions about what constitutes "good" higher order thinking and communication can be based on knowledge of a wide range of styles from various groups. In other words, perhaps the "best" can be chosen from a variety of cultural traditions to improve communication in educational and business/professional contexts (Kochman, 1992).

In order to clarify the problems involved in cross-cultural and cross-gender communication, this brief response paper will specify a few illustrative examples of both culture- and gender-based differences in communicative style. These examples are taken from a substantial body of research, primarily from the disciplines of linguistics and anthropology. Before doing so, however, it is important to define some basic concepts upon which the research is based that should be useful in any assessment of communication.

Hymes' (1972) notion of communicative competence is essential in this discussion. Such competence, according to Hymes, is "dependent upon both (tacit) knowledge and (ability for) use" (p. 282) and includes all modes of communication, verbal and nonverbal. "The requisite knowledge includes not only rules for communication (both linguistic and sociolinguistic) and shared rules for interaction, but also the cultural rules and knowledge that are the basis for the context and content of communicative events and interaction processes" (Saville-Troike, 1983, pp. 2-3). The fact that both linguistic (e.g., pronunciation and grammar) and sociolinguistic (e.g., what sentences are appropriate in what contexts) rules are a part of communicative competence underscores the culturally defined nature of interaction. That is, the way a person goes about communicating with other people calls upon deeply ingrained and mostly unconscious knowledge acquired as part of that person's socialization within his or her particular cultural group. Because these patterns are both deeply ingrained and largely unconscious, communication across cultures (and genders) can be quite problematic (Gumperz, 1982 a, b; Tannen, 1990). Such problems arise frequently in both the workplace and in educational (and other) institutions.

Another term needs attention (defining it is far beyond the scope of this brief response paper). Literacy, for the purposes of the National Adult Literacy Survey, includes "using printed and written information" but excludes "teamwork skills, interpersonal skills, and communication skills." Daly criticizes this, citing Ong (1982). Actually, Ong is a leading theorist in dichotomizing orality and literacy, not conflating them. I have argued (Farr, in press), as has Graff (1981), that the only definition of literacy that extends across time and space (and thus throughout history and across cultures) is one similar to that being used by the NALS. While many studies (e.g., Finnegan, 1988) have shown that "higher level" thinking occurs in both oral and literate modes, the two should not be conflated, as they are not "the same" (e.g., Goody, 1986, 1987; Finnegan, 1988). Instead, both orality and literacy should be considered, and in fact are, of value. And it is important to note that the communicative styles into which one is socialized are expressed in both oral and written modes of language. Thus, the way one speaks can often be identified in the way one writes.

A final concept is important here: variation. A cultural pattern is not generalizable to an entire group. That is, individuals within any group (be it based on culture/ethnicity or gender) vary along a number of continua. To say, then, that a communicative style is associated with a particular group is not to say that all members of that group share it equally. All formally educated people, in fact, have acculturated to some degree to the communicative style, often called "essayist literacy", which has evolved historically with Western societies (Farr, 1993). This means that, although cultural background and gender must be taken into account in assessments of communicative abilities, it cannot be assumed that all females or members of ethnic minorities always communicate in the style generally associated with their group. All people shift communicative styles across contexts, and many people are bicultural and bilingual or bidialectal.

The studies cited in Daly's paper seem primarily to have been carried out with white, culturally "mainstream" subjects, or at least are based on the communicative norms of that population. If the subjects in these studies were not of such homogeneity, then perhaps that explains his comment that "many people are not that effective" at communicating (p. 4). In these studies, could negative evaluations of effectiveness be explained by differences in gender? In cultural background? If "study after study suggests that in work situations employees sorely lack effective speaking and listening skills" (p. 4), what kind of employees were these? What kinds of speaking and listening? In what contexts? Although Witte's paper emphasizes the cruciality of context and proposes an event-centered assessment process, he also omits consideration of the effect of cultural background and gender on communication.

A few examples will illustrate the relevance of cultural and gender differences to assessment criteria. Appendix F in Daly's paper, a rating form for "conversational skills," lists behaviors that are classically defined in cultural terms: use of eye contact, interruption of partner, speaking rate, lean toward partner (neither too far nor too far back), speaking about self, use of humor, vocal volume, expression of personal opinions, facial expressiveness, use of gestures, etc.

Scollon and Scollon (1981) traced problematic communication between Athabaskans in Alaska and North American English speakers to a difference in pause length between conversational turns. Because Athabaskan communicative conventions mandate a slightly longer pause between speakers than do North American "mainstream" conventions, English speakers tend to do most of the talking in cross-cultural conversations. When English speakers stop and wait the length of time they expect to wait before the other speaker begins speaking, they encounter silence (because Athabaskans are waiting slightly longer before taking their turn at speaking). As a consequence, English speakers are seen by Athabaskans as egotistic and verbally domineering, and Athabaskans are seen by English speakers as taciturn and uncooperative. The Athabaskans say they "can't get a word in edgewise" and the English speakers say they "can't get the Athabaskans to say anything!"

Kochman (1981) describes cultural differences between African Americans and whites in what is considered persuasive communication. At a public meeting of white academics and black community activists in Chicago, communication was problematic precisely because style of arguing differed. The whites thought that they couldn't discuss the differences that the meeting was called to resolve because the blacks were "too emotional"; they suggested waiting to discuss the differences later, when the others had "calmed down." The black activists, in contrast, considered it quite appropriate to argue with emotion and vehemence; to them, such behavior reflected commitment and honesty. When the whites at the meeting argued dispassionately, according to their own norms of "effective" speaking, they were considered hypocritical by the blacks, who judged them, as everyone does, by their own communicative norms.

In my own study of communicative norms in the Mexican-origin community in Chicago (Farr, 1993), Mexican immigrants display a very different rhetoric in public speaking than do members of other groups, including Mexican Americans. People enculturated in Mexico weave an elaborate, and poetic, "tapestry" with their words, and when they use this style to argue in public meetings, they are sometimes, depending on the context (and, crucially, their audience), interrupted impatiently by those who prefer concise statements of particular "points," values associated with a U.S. "mainstream" style of communication and reflected in the way U.S. educational institutions teach writing.

Tannen (1990) provides numerous examples of gender-based differences in communication. Synthesizing a body of research on gender differences in communication, Tannen argues that men, in general, approach communication in a more hierarchical, competitive way than do most women, who generally approach communication in a more collaborative manner. Cautioning that this does not mean that all men and women use these gender-associated styles (individual as well as cultural differences play a role here; see Kochman (1981) regarding the relatively assertive style of many African-American women), she nevertheless illustrates, with evidence from cross-gender communication studies, how communicative difficulties can ensue when styles differ starkly. Although Tannen provides numerous examples of male-female interpersonal communication, she does not provide examples of such communication in the workplace. Her synthesis of research, however, can be extrapolated to a workplace situation. Many women, for example, tend to approach communication in the workplace as a collaborative venture, often expecting that the opinions and comments that they offer will be part of a dialogue in which a decision will be reached. Many men, in contrast, approach communication in the workplace as an opportunity for a competitive verbal display of their own ideas and opinions, and expect others to compete similarly. Such a conflict in expectations regarding the purpose of communication clearly leads to frustrations and, often, miscommunication.

Many more examples could be cited, especially those relating to references in Daly's paper to "appropriate turntaking," "appropriate opening and leavetaking" (p. 21), and "interrupting" (p. 11), as well as to "politeness" (p. 26), which is quintessentially culturally defined (what is polite in one culture is rude in another--politeness, like humor, does not "translate" well). I hope, however, that the present examples suffice to emphasize the cruciality of cultural and gender differences. The argument might be made by some that college graduates should acquire the traditional (educated, male, European) communicative norms, so such differences don't really matter. But they do matter, especially in colleges and universities where women and minorities often have difficulty "getting into the conversation" and thus acquiring such skills. If we don't take such differences into account, in both teaching and assessment, we face the probability of continued problems in the "educational preparedness" of college graduates and future members of the workforce. At the very least, we must be certain that what we do decide to measure is actually important for effective communication and "higher order" thinking in specific contexts, and not simply what traditionally has been perceived as important by one particular group of people.

References

Farr, M. (in press). Biliteracy in the home: Practices among Mexicano families in Chicago. In D. Spener (Ed.), *Biliteracy in theory and practice*. Washington, D.C.: Center for Applied Linguistics.

Farr, M. (1993). Essayist literacy and other verbal performances. *Written Communication*, 10 (1).

Finnegan, R. (1988). *Literacy and orality*. New York: Basil Blackwell.

Goody, J. (1986). *The logic of writing and the organization of society*. New York: Cambridge University Press.

Goody, J. (1987). *The interface between the written and the oral*. New York: Cambridge University Press.

Graff, H. (Ed.). (1981). *Literacy and social development in the West*. New York: Cambridge University Press.

Gumperz, J.J. (1982a). *Discourse strategies*. New York: Cambridge University Press.

Gumperz, J.J. (Ed.). (1982b). *Language and social identity*. New York: Cambridge University Press.

Hymes, D. (1972). On communicative competence. In J.P. Pride & J. Holmes (Eds.), *Sociolinguistics*. New York: Penguin Books.

Kochman, T. (1981). *Black and white styles in conflict*. Chicago: University of Chicago Press.

Kochman, T. (1992). *Effective management of cultural diversity, a seminar series*. Chicago: Kochman Communication Consultants.

Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. New York: Methuen.

Scollon, R., & Scollon, S. B. (1981). *Narrative, literacy and face in interethnic communication*. Norwood, NJ: Ablex.

Saville-Troike, M. (1983). *The ethnography of communication: An introduction*. New York: Basil Blackwell.

Sima, C. (1987). *The Hispanic community of the future: Implications for UIC student enrollment*. Strategic planning brief, vol. 1, no. 5. Chicago: University of Illinois at Chicago.

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York: Morrow.

Part 2. THE WORKING GROUPS

Introduction to Working Group Reports

After experiencing the enormous range of debate that developed around the first November conference a year earlier (see citations noted in the first section), NCES planners were looking for a way to harness the energies of those attending--only a few of whom had participated in the first conference--to a fairly focused task: try to develop a working consensus on the set of critical thinking, problem solving, and communication skills needed by college graduates. This target is easier to state than to hit.

To mention only one confounding factor, the prospect of assessing higher order abilities on such a national scope provokes (among people familiar with the terrain) a tendency to issue dire warnings and proclamations of doom. As most of the scholars in both workshops were chosen because of their background either in the domains under consideration or in assessment (or both), they were in an excellent position to appreciate the enormity of the task of "inventing" a national assessment of college student learning. Over a third of last year's conference product was devoted to articulating some of these threshold issues, major caveats and concerns. But last year's group *was able*, nonetheless, to begin to suggest how to identify and catalog the skills involved. It was upon this foundation that the second conference was to stand, permitting a more pragmatic focus on the nature of the skills themselves.

Each of the four groups met in working sessions that ran to nearly 10 hours. They shared their deliberations with one another only informally, though especially at the closing plenary session. As mentioned earlier, each group was encouraged to adopt the icon as a meeting tool and organizing format. As the separate working sessions ensued, each began to take a fairly distinct form, and while group I and to a lesser extent group II were able to employ the icon format usefully, the group III and especially the group IV discussions found it somewhat counterproductive, and discovered their own way of organizing fruitful discussion.

The group summaries that follow were reproduced in retrospect. All sessions were recorded by machine, and by recorders who worked with the facilitators to try to develop an accurate record of the group's conversation. From these sources, and from summary charts created by the group, the following reports were derived by Addison Greenwood, the author of some of the previous workshop's results. The advantage of such a method is to impose a somewhat coherent view of what transpired. This necessitates filtering the original source material through one interpreting source, whose consistent goal was to elicit from the many observations and comments those most germane to the target specified by NCES, namely a listing of skills. The corollary sacrifice of reporting the richness, diversity, and the many fine points of discussion is a necessary function of moving such complex and subtle discourse toward the realm of political recommendations. The summaries that appear here have been vetted by the authors and facilitators however, and it is hoped they capture the spirit of the four groups' deliberations.

One other bit of circumstantial history is appropriate. A special thanks should go to Jeff Gilmore, Addison Greenwood, and Angie Miles. They took charge and conducted the workshop on short notice, due to an unexpected detachment of the retina of the right eye of Sal Corrallo, project director and workshop coordinator.

PROBLEM SOLVING AND CRITICAL THINKING, GROUP ONE - AUTHOR - HALPERN

PARTICIPANTS:

Facilitator:

Jeffrey Gilmore
OERI, Office of Research

Author:

Diane Halpern
California State University, San Bernadino, Dept. of Psychology

Reviewers:

Michael Mumford
George Mason University, Center for Behavioral and Cognitive Studies

Richard Paul
Sonoma State University, Center for Critical Thinking

Mary Tenopyr
AT&T, Selection and Testing

Participants:

Clyde Ingle
Commissioner, Indiana Commission on Higher Education

Nancy Kari
College of St. Catherine, Professor

Karen Martinez
American Assembly of Collegiate Schools of Business, Director of Accreditation

Mary Anne Nestor
Office of Personnel Management, Professional and Administrative Examination Branch

James Ratcliff
Pennsylvania State University, National Center on Postsecondary Teaching, Learning, and Assessment

Ronald Thornton
Tufts University, Center for Science and Math Teaching

Shari Tishman
Harvard University, Project Zero

Alexandra Wigdor
National Academy of Sciences, Division on Education, Training, and Employment

Joan Wills
Institute for Educational Leadership, Center for Workforce Development

Others:

Mary Anne Statman
Recorder, Consultant, Washington, DC

Sheila Maramack
OERI, Office of Research

John Mathews
OERI, National Center for Education Statistics

Nevzer Stacey
OERI, Office of Research

PROCEEDINGS:

Group One facilitator Gilmore was the most successful at marshaling his group's product into this format. "As it turned out, the group actually did this five times. That is, after poring over, discussing, and revising a partially filled out Icon Chart prepared by the facilitator (based on the authors' papers), the group decided to develop four additional, separate conceptualizations of critical thinking/problem solving based on the papers by Halpern, Mumford, Chaffee, and Perkins/Tishman."

Here follows an attempt to coalesce the insights from the three systems whose creators participated in the Group 1 discussion: author Diane Halpern and reviewers Michael

Mumford and John Chaffee. As might be expected in trying to shoehorn such distinct and rich collections of ideas into a catalog, there arise a number of problems of method: Is one scholar's disposition another's cognitive frame? Are similar underlying cognitive skills intended though masked by the choice of different terms? Some overlap, redundancy, and even perhaps some mismatching is unavoidable. But the group's effort to approach each of these rich and diverse systems in a consistent way may be the first step in trying to reconcile their differences in aid of developing a collective model. The group did not determine that such a hybrid was preferable to the alternative of testing and developing multiple models, and thus considered each model on its own terms and merits. But it may come about that no one single critical thinking model will emerge as superior to all the others, and this beginning of a search for consensus may be useful.

Critical Thinking: Meaning, Unique Aspects

A first proviso: As was the case in the first conference, no major distinct category of skills or thinking was attributed to the domain of "problem solving," as such. While there may of course be distinct ways of framing problems according to one or another venerable systems of analysis—doubtless going back to the ancient Greek arts of rhetoric—the insights developed during earlier work about the domains involved in Goal 5 were reinforced here. Critical thinking is what people should do with problems in their cognitive experience (consider this one large domain), and thus the separate terms originally invoked by the Goals Panel should *not* suggest the existence of distinct domains of critical thinking skills for each, or distinguishable ways to solve problems other than by critical thinking. Critical thinking is an activity people engage in to—among other things—solve problems. For the purposes of the assessment exercise NCES is exploring, it was almost universally felt, the terms should be collapsed: if not critical thinking, perhaps higher order thinking.

Critical thinking/problem solving is:

- Purposeful and self regulatory judgement reached by exercising generic cognitive skills in analysis, interpretation, inference, evaluation, and explanation.

- Characterized by the active construction of knowledge; by knowing how to learn and adapt; by the ability to solve new problems, and by the willingness to change basic concepts and beliefs.

- The successful use of particular skills (e.g., skill(s) in applying, analyzing, synthesizing, and evaluating information) in an appropriate context. It is also an attitude or disposition to recognize when these skills are needed, and the willingness to use and apply them.

- A habit of mind that one develops: to be self-critical.

It should be emphasized that critical thinking/problem solving is *not*:

- The simple acquisition of a discrete set of thinking and language "tools." Instead, it involves developing an integrated set of thinking abilities, language constructions, critical attitudes and fundamental beliefs that involve the whole person. It is a transformational process that involves a total approach to making sense of experience, changing the way that people view the world and make decisions.

- Speculation, meditation, contemplation, or appreciation.

- Limited to mastery of content areas; it has transferability.

At the collegiate level, **critical thinking skills**:

- Are the ones that are promoted in thoughtful college classrooms in which students grapple with complex issues and problems through discussions, texts, assignments, examinations, and activities.

- Are often referred to as "higher order cognitive skills" to differentiate them from simpler (i.e., lower order) thinking skills. Higher order skills are relatively complex, require judgment, analysis, and synthesis, and are not applied in a rote or mechanical manner. Higher order thinking is thinking that is reflective, sensitive to the context, and monitored.

Crucial Issues of Classification

A second proviso: No clear consensus exists about the use of certain terms which recur throughout the discussion of critical thinking in the Academy. Again, this phenomenon echoes the experience of the first conference, where certain common definitions were established de facto for the purposes of discussion. This is not intended to diminish the *importance* of such distinctions and categories, only to note that one person's genre is another's domain, is another's dimension. What the group found most important was how the underlying assumptions of a given system or taxonomy related to the eventual classification and division of elements in that particular system, whether those elements are referred to as macroskills, microskills, competencies, abilities or dispositions.

As clarified in her paper (and echoing many from the first conference), Halpern makes one major distinction: between the abilities one must have mastered, and the dispositions one manifests in order to display those abilities. She breaks each of these two realms down into five areas. Mumford's Construct List provides five realms, three of which are referred to as species of *Dimensions*, two as species of *Skills*. The group believed that his *Creative Process Dimensions* were in fact abilities, more or less in the sense Halpern uses the term, while the Dimensions labeled Wisdom and Adaptability tended to fit her Dispositions category. Chaffee puts a great emphasis on the process of developing as a critical thinker, and on the necessity to evaluate such thinking only in the context of tasks that occur in "real-life." Thus, his text is framed in terms of competency objectives, which cannot fairly be considered by trying to first classify and then disentangle abilities and dispositions. For the other two, this distinction seems useful:

Thus, the largest breakdown of sets of abilities can be seen as:

> For Halpern (5): Verbal Reasoning Skills, Argument Analysis Skills, Skills in Thinking as Hypothesis Testing, Using likelihood and Uncertainty, and Decision Making and Problem Solving Skills, and
>
> For Mumford (3): Interpersonal skills and Cognitive Processing, which includes Creative dimensions as well as Practical skills.

The largest breakdown of sets of dispositions can be seen as:

> For Halpern (5): Willingness to engage in and persist at a complex task; Willingness to plan; Flexibility; Willingness to self-correct; Being mindful; and
>
> For Mumford (2): the Dimensions of Adaptability and Wisdom.

Distinguishable Abilities

Both Halpern and Mumford provide a breakdown of their larger sets into more discrete abilities and dispositions. First, a summary of the lists and Chaffee's catalog, and then some observations about all three taxonomies.

Halpern

Verbal Reasoning Skills

a. recognizing and defending against the inappropriate use of emotional and misleading language (e.g.,labeling, name calling, ambiguity, vagueness, hedging, euphemism, bureaucratese, and arguments by etymology [original word use]);

b. detecting the misuse of definitions and reification;

c. understanding the use of framing with leading questions, negation, and marked words to bias the reader;

d. using analogies appropriately, which includes examining the nature of the similarity relationship and its connection to the conclusion;

e. employing questioning and paraphrase as a skill for comprehension of text and oral language (i.e., recognizing main ideas); and

f. producing and using a graphic representation of information provided in prose form.

Argument Analysis Skills

a. identifying premises (reasons), counterarguments, and conclusions;

b. reasoning with "if, then" statements (which includes avoiding the fallacies of affirming the consequence and denying the antecedent);

c. judging the credibility of an information source;

d. judging the consistency, relevance to the conclusion, and adequacy in the way premises support a conclusion;

e. understanding the differences among opinion, reasoned judgment, and fact; and

f. recognizing and avoiding common fallacies such as straw person, appeals to ignorance, slippery slope, false dichotomy, guilt by association, and arguments against the person.

Skills in Thinking as Hypothesis Testing

a. recognizing the need for and using operational definitions;

b. understanding the need to isolate and control variables in order to make strong causal claims;

c. checking for adequate sample size and possible bias in sampling when a generalization is made;

d. being able to describe the relationship between any two variables as positive, negative, or unrelated;

e. understanding the limits of correlation reasoning;

f. seeking converging evidence to increase confidence in a conclusion;

g. considering the relative "badness" of different sorts of errors;

h. solving problems with proportional and combinatorial (systematic combinations) reasoning; and

i. determining how self-fulfilling prophecies could be responsible for experimental results and everyday observations.

Using Likelihood and Uncertainty

a. recognizing regression to the mean;

b. understanding and avoiding conjunction errors;

c. utilizing base rates to make predictions;

d. understanding the limits of extrapolation; and

e. adjusting risk assessments to account for the cumulative nature of probabilistic events.

Decision Making and Problem Solving Skills

a. listing alternatives and considering the pros and cons of each;

b. restating the problem to consider different sorts of alternatives;

c. recognizing the bias in hindsight analyses;

d. seeking information to reduce uncertainty;

e. recognizing decisions based on entrapment;

f. producing graphs, diagrams, hierarchical trees, matrices, and models as solutions aids;

g. understanding how world views can constrain the problem solving process; and

h. using numerous strategies in solving problems including means-ends analysis, working backward, simplification, analogies, brain storming, contradiction, and trial and error.

Mumford

Mumford's Construct List was included in his review. The group generally agreed with its coherence, but modified some of the entries to better reflect the breakdown into abilities/dispositions. (Those elements of his taxonomy thought to fall outside this particular analysis appear in parentheses.)

Creative Process Dimensions

a. Problem Construction/Problem Finding

b. Information Encoding

c. Category Search

d. Specification of Best-Fitting Category

e. Category Combination and Reorganization

f. Solution Monitoring

(Idea Evaluation)
(Solution Implementation)

Practical Processing Skills

a. Prioritizing

b. Time Management

c. Problem Anticipation

d. Information Gathering

e. Evaluation of Discrepancy Importance

f. Monitoring and Implementation of Solution Outcomes

g. Knowing When to Act

(Information Appraisal)
(Implementation of Solutions)
(Planning and Implementation)

(Monitoring and Assessment of Goal-Relative Cues)
(Selection of Solution Components and Information)

Interpersonal Skills

a. Negotiating Skills

b. Social Sensitivity

c. Empathy

d. Behavioral Flexibility

(Social Confidence)
(Persuasion)
(Emotional Control)

Distinguishable Dispositions

Halpern

Critical thinking is more than the successful use of a particular skill in an appropriate context. It is also an attitude or disposition to recognize when a skill is needed and the willingness to apply it. There is an important distinction between what people can do and what they actually do in real world contexts. This is called the competence-performance distinction. It is of no value to teach students the skills of critical thinking if they don't use them. A critical thinker will exhibit the following dispositions or attitudes:

1. Willingness to engage in and persist at a complex task

Critical thinking is hard work that requires diligent persistence, a trait that is central to academic success and most other success in life. In the jargon of cognitive psychology, tasks that require critical thinking have a high mental workload. A critical thinker must be willing to expend the mental energy that is required to begin and complete the task.

2. Willingness to plan

Psychologists call this the ability to inhibit action. A planful approach requires individuals to check their impulsivity and engage in intermediate tasks such as seeking additional information, organizing facts, and generating alternatives. The willingness to plan must become a habitual approach that is applied in many different contexts.

3. Flexibility

An attitude of flexibility is marked with the willingness to consider new options, try things a new way, reconsider old problems. It is the antithesis of the rigidity and dogmatism that is characteristic of a "closed mind."

4. Willingness to self-correct

This refers to the willingness to learn from errors instead of becoming defensive about them. Individuals with the willingness to self-correct are able to utilize feedback and recognize the factors that led to the error. In order to improve the thinking process, ineffective strategies and automatic responses need to be recognized and abandoned.

5. Being mindful

Psychologists call this trait metacognition or metacognitive monitoring. It is the tendency to monitor one's comprehension and progress toward a goal. Critical thinkers develop the habit of self-conscious concern for and evaluation of the thinking process.

Mumford

Adaptability Dimensions

a. Creative Achievement

b. Self-Discipline

c. Environmental Engagement

d. Defensive Rigidity (Negative Influence)

(Also Negative Influences:
Interpersonal Competitiveness and Evaluation Apprehension)

Wisdom Dimensions

a. Self-Reflectivity

b. Social Perceptiveness

c. Sensitivity to Solution Fit

d. Practical Intelligence

(Diversity of Experience)
(Moral Awareness)

In his review of Halpern's paper, Mumford emphasizes "the need for a clear-cut theoretical model" so as to obtain construct validity and to provide "a foundation for stable scientific conclusions." Had this been done with the Halpern taxonomy, he believes, it would show an undue focus "on skills required for critical thinking in academic settings, especially in the sciences and social sciences." He specifically cited several of the skills he listed under *Creative Process Dimensions* as missing from Halpern's catalog, suggesting her analysis may not "adequately cover complex creative-thinking skills." He also believes that many of the skills he categorized as *Practical Processing* and *Interpersonal* need to be added in order to reflect "certain basic cognitive skills found in the workplace," many of "which are of substantial importance in jobs emphasizing teamwork." He appreciates that, "in making these revisions, it would be desirable to formulate a more integrated conceptual approach that links workplace and learning issues and provides adequate substantive or empirical justification for the proposed taxonomic categories."

Critical Thinking at LaGuardia College

Chaffee's approach to critical thinking [elaborately developed and described in his textbook *Thinking Critically* (3/E)] grows out of a course that has been evolving under his directorship of Creative and Critical Thinking Studies at LaGuardia College, City University of New York, since 1979. A first premise is that "thinking is a process that can be understood and improved through proper study and practice." A second course has been developed in due time entitled *Creative Thinking: Theory and Practice*, which "focuses on the cognitive processes we use to generate and refine innovative ideas." This emphasis echoes Mumford's cognitive lens.

Where the LaGuardia program seems to demand a distinct appraisal is from the "conviction that thinking and literacy abilities can only be taught effectively through a process of synthesis, giving students the means to clarify and make sense of themselves and the world in which they live." This drives the model toward assessing only "challenging, complex, and clearly structured performance tasks that are accompanied by explicit criteria." Chaffee's system assumes the inherently complex nature of all knowledge and thus the continuously ongoing synthesis critical thinkers are engaged in constructing. Critical thinking thus is grounded in meta-critical thinking, which takes the form of a reflective analysis of the processes of thinking, language, and knowing.

As the LaGuardia model is detailed in Chaffee's text, it reveals many of the elements found in the other two systems, though Chaffee's emphasis on a developmental/epistomological framework precludes sharp distinctions between dispositions and abilities, and virtually forbids any listing or cataloging of discrete abilities outside the context of "the complex interactions in which they typically occur." Following the sequence of chapters in the text, many of the topics cover basic elements of the other systems, though the emphasis on task-oriented real-world contexts continually interweaves dispositions with abilities.

Thinking for Chaffee begins with *the understanding that it is an active, purposeful and organized process that is used to make sense of the world.* This takes several major forms which require related subskills:

a. **Achieving goals**: identify appropriate goals, devise effective strategies for achieving them.

b. **Analyzing complex issues:** identify the issue, describe multiple interpretations of it, identify and evaluate the evidence and arguments to support these various interpretations, and reason through to and articulate an informed conclusion.

c. **Making decisions**.

Critical Thinking *involves two major ways of thinking and several strategies, which serve to make sense of the world by carefully examining the thinking process to clarify and improve understanding.* One must be:

a. **Thinking actively**: Take initiative, follow through, take responsibility, and be constructive/productive.

b. **Thinking autonomously:** Develop informed, well-reasoned conclusions that draw on the views of others, but which represent one's own independent analysis/synthesis and conclusions, supplying necessary reasons/evidence, authorities, references, factual evidence, and personal experience.

Thinking critically in these ways, one must then:

c. **Carefully explore situations with questions:** Analyze complex situations with questions at all cognitive levels (i.e., fact, interpretation, analysis, synthesis, evaluation and application).

d. **View situations from different perspectives:** Get outside one's own frame of reference, and think empathetically within other (contrasting and conflicting) points of view, supporting diverse perspectives with reasons/evidence.

e. **Discuss ideas in an organized way:** Engage in a dialectical exchange of ideas through dialogue by listening, supporting views, responding to points being made, asking and trying to answer important questions, and trying to increase understanding rather than just winning an argument.

Problem Solving *involves the application of an organized approach to solving complex, open-ended problems.* One must ask:

a. **What is the problem:** Seek knowledge, results, and a definition.

b. **What are the alternatives:** Identify the boundaries and alternatives.

c. **For each alternative, what are the dis/advantages:** Gather information.

d. **What is the solution:** Develop alternatives and a plan of action.

Perceiving *involves awareness and evaluation.* One must understand:

a. **The active nature of the perceiving process:** Be aware of actively selecting, organizing, and interpreting sensations in ways that reflect one's individual perceiving "lenses."

b. **And then critically evaluate how these lenses are formed through experience and reflection.**

Believing and Knowing *also involves awareness and evaluation.* One must understand:

a. **The way beliefs function in making sense of one's experience to interpret, evaluate, conclude and predict.**

b. **The interaction between the processes of perceiving, believing, and knowing.**

c. **The process by which beliefs are formed and revised, based on analysis and critical evaluation.**

One must evaluate:

d. **The accuracy of beliefs:** Understand and be able to apply various criteria by providing compelling and coherent explanations; determine consistency with other well-supported beliefs; provide accurate predictions about what events will occur; support with substantive reasons and evidence derived from reliable sources; demonstrate their falsifiable nature; i.e., state the conditions (tests) under which the beliefs could be disproved, and put the beliefs through those tests.

e. **The reliability of information and its sources.**

Mastering Language as a System *involves mastering the way language functions as a system of symbols for thinking and communicating.* One must understand (and, when required, be able to interpret/construct):

a. **The multiple levels of word sense:** Appreciate the various kinds of meaning, such as the semantic, perceptual, syntactic, and pragmatic.

b. **Various forms of sentence meaning:** Appreciate the various kinds of sentences, such as simple, complex, and coordinated.

c. **The relationships between and among the sense(s) of words, the meaning of sentences, and the process of thinking.**

d. **The different ways language is used in various social contexts:** Distinguish and appreciate language styles, slang, jargon, and dialects.

e. **The way language is used to influence the thinking, feelings, and behavior of others:** Realize the effects of euphemistic and emotive language and the language of advertising.

One must be able to:

f. **Use language as a tool:** Be clear and specific, and avoid vagueness, ambiguity and cliches.

Forming and Applying Concepts. One must understand the:

a. **Structure of concepts and how they serve to identify and organize experience:** Distinguish among signs, their properties and their referents.

b. **Process of forming concepts through the interactive activities of generalizing and interpreting.**

c. **Process of defining and applying concepts.**

Relating and Organizing. One must:

a. **Create mind maps for thinking and using language.**

b. **Understand, analyze, and create thinking patterns to organize experience and to make sense of the world:** Learn the intricacies of the various types of relationship; that is, causal, comparative and analogical, and chronological and process relationships.

Reporting, Inferring, and Judging *involves understanding, analyzing and using judgment, reason and logic to evaluate information, inferences, and conclusions for accuracy, consistency, appropriateness, and other relevant attributes.*

Constructing Arguments. One must be able to:

Recognize an argument and understand its function and structure: Recognize the basic deductive form, and the reasons and conclusions of an argument, and evaluate an argument for truth, validity, and soundness.

Reasoning Critically *involves mastering the forms of inductive reasoning.* One must be able to:

a. **Evaluate their plausibility:** Recognize common fallacies, and those associated with certain forms of inductive reasoning.

b. **Use them effectively:** Develop empirical generalizations, appreciating whether the sample evidence is known, valid, sufficient and representative, and whether an experiment is properly controlled; and be adept at reasoning from causes and other tools of the scientific method.

Discussion

Once agreement was reached to consider each framework separately, Group 1 was able to proceed to an energetic and productive discussion of the many merits and aspects each in turn offered. This same process was also unfolding down the hall in Group 2 [See next report]. The glass-half-empty way of looking at this drama suggests that—in these two groups—at least six scholar/team(s) came to the workshop to promote and defend their own system and taxonomy, and the six went away bloodied but not bowed, with no clear "winners or losers." As Chaffee put it, each of the system/proponents has a common goal: to carve up reality coherently and comprehensively; in philosophical terms, to define the problem space and to propose a solution. But because the focus and underlying assumptions of each is unique (Mumford emphasized), it is not simply a matter of looking at one system for "items" (skills) that are missing from another and combining them together into a superlist. As the group saw it, trying to map one onto another is a task with daunting philosophical hurdles to overcome.

Nonetheless, the papers included in this volume constitute a possible first step in that direction. Published here are the authors' *revised* papers. Considering all of the criticism offered by reviewers to their first draft, as well as their group's two days' worth of examination and analysis, the authors for this present publication were able to revise their papers where they thought appropriate. Whether any of these revisions constitute a sufficiently ecumenical evolution to reflect all possible improvements toward a consensus system that could serve as a template for the "final" version remains to be seen.

Ratcliff suggested, however, and many in the group concurred, that the glass was actually half full. While his remarks were made in response to the question of what the Center for Postsecondary Teaching, Learning and Assessment at Penn State might actually *do* with all of this material [see concluding section of this report], the group generally ratified the concept that such "contending" systems might—in a delphi or some other evaluative process—be compared to one another according to some overarching criteria. [Clearly, most of the authors believe their system complete, coherent, and philosophically sound, and thus a comparison of their respective underlying theoretical frameworks and "ways of carving up reality" might just invite more debaters onto the head of a pin.] But what might such criteria be?

That question may turn out to be important to the Penn State task of creating the survey instrument itself, and could even find its way into the *content* of the survey, if it seems crucial to successfully taking the national assessment of college student learning to the next level. Group 1 thought several criteria should apply to such a comparative analysis of systems and taxonomies. First and most important: Are the thinking skills comprehensive and are they learnable and teachable? Second: Can they be embedded into the curriculum so as to enhance transfer and to assure that students are receiving a coherent instruction within and across disciplines, although there is an inherent problem that the heritage implicit in certain traditional modes of inquiry could be undermined? Third: Can the entire system produce results, however that comes to be defined; that is, link educational practices and outcomes? The group believed this was just a good first guess at what these framework evaluations might include, but hoped the value of maintaining multiple models could persist through as much of the national assessment development process as was constructive.

Another set of concerns was voiced with respect to the conceptual core of critical thinking as it was being developed, and to how it might make the translation through the assessment development process, onto the campus, and eventually to be tested as a lens on critical thinking/problem solving in the arenas specified by Goal 5, the workplace and citizenship contexts. [These concerns, to varying degrees, were to be echoed by the other groups as well.]

○ **Diversity of students versus developing an instrument that will provide a fairly muscular comparative framework.** Students across and within institutions present a wide array of backgrounds and ability levels. At some fairly central level, the national assessment needs to decide whether this diversity is to drive the form of the assessment, or whether the goal of the process is to provide a very utilitarian tool to assess abilities and success by some standard. A tilt toward the latter would be difficult to do in the light of the realization that demonstrating so many of these skills is culturally bound. While there may be a universal, cognitive core to the target, teasing it into a valid instrument or system is the real challenge. Given the existence of various instruments, the group thought it much more likely that a menu of mixed measures—which could be adapted within and across various groups—would succeed.

○ **Making present goals explicit while emphasizing the ubiquitous and continuous domain of higher order thinking.** The group was nearly unanimous in their concern about the underlying conceptual integrity of higher order thinking. Must it be democratized—in order to meet some of the concerns in the previous bullet—to the point where it no longer served to truly point to the standard originally envisioned by the Goals? A first premise developed by the group was that higher order thinking is a continuous process, evident not only in class but throughout life, both at the university and continuously after graduation. They believed that a better definition of the scope of the present exercise should reflect a very utilitarian view, driven by the best that all parties to the assessment thought practicable. Wherever problems can be anticipated, whether with faculty or other institutional elements, or with particular student populations, planners should try to build a sufficient flexibility into the process in order to keep moving forward. A major issue is how to embed the skills into the current educational system, avoiding the specter of an expensive add-on that could be undermined.

○ **Better focus of the entire enterprise, and better communication with the workplace constituency.** As an example, group member Joan Wills described the results of extensive work with industry done by the Center for Workforce Development. This would be one component to assure that the assessment goals and skills remain true to the competitiveness aspect of Goal 5. Another is to coordinate the various government efforts at diverse agencies under one overarching structure. Finally, identify all of the key players and try for a consensus through pilot testing and multiple measures.

The "Penn State" Process

The NCES plan to use the results of the present conference to inform a survey instrument of some 600 interested parties was explained at the outset of the conference. Though not part of the workshop's original agenda, members of all four groups were intensely interested—many expressing concern—in how this process would work. Group 2 developed a couple of "products" to accompany the group reports to the National Center at Penn State. Group 1 had the benefit of the director of this next step, James Ratcliff of the National Center for Postsecondary Teaching, Learning and Assessment (NCTLA), joining in their discussion. Here follows conference host Jeff Gilmore's remarks by Ratcliff about the process, and an excerpted version of remarks made throughout the Group 1 meetings about what everyone was calling "the next step" in the national assessment of college student learning development process.

JEFF GILMORE in the Opening Session:

As Gary and Emerson pointed out, this conference is not an isolated event. This is one in a series of events, starting last year with our conceptual conference on assessing, and now we're going to get to work and come up with some specific competency areas and skills. The next step is that the products from this conference, which will be primarily the authors' papers, the reviews, the revised authors' papers that come out of the conference conversation, and what we will call the ICON chart, and more about that in a minute.

These will be transmitted over to the NCTLA at Penn State for the next step. And what that next step will be, very briefly, is that the National Center for Postsecondary Teaching, Learning and Assessment will take these materials. They will begin their own literature reviews; they've done quite a bit of that already. They'll be contacting the appropriate education, cognitive, and area disciplinary associations. And, building on their own research of the last few years, coming up with an instrument they will then send out in a delphi technique, in two iterative rounds, to 600 faculty members, policymakers, and business people, to try to arrive at some further consensus building on the specific skills that we would wish to assess.

JAMES RATCLIFF, during Group 1 discussions:

Goals of the Exercise

First off, these are baby steps. . . And where we're going with this project in the next six months—as I said it's a very modest project—to be able to say that faculty, employers and government officials associated with institutions that teach a nationally representative sample of students feel that these criteria and these typologies do in these ways—and don't in these ways—represent what they're currently doing. This is what NCES would like—and this is a project that has been defined largely by what NCES would like—as the next formative step. We at the National Center would take that step for them. They would like some answers to the validity of the constructs that we discuss here on campus, and to determine if there is some consensus about the validity of these constructs. Validity within the context of current practice, that is.

This is not to necessarily say that what we've worked on here should be modified by the information from the survey. It's rather an attempt to contrast a discussion of what *should be* Goal 5, Objective 5 criteria with a discussion of at least what people intend to do in current practice relative to those criteria. Thus, it's not our intent to necessarily try to create mashed potatoes out of this stuff, but rather to have a description of current practice relative to proposed criteria.

On the Role of Faculty in the Assessment Development

I think there are several levels of validation that are required in order to make a national assessment effective. It is not our intent to touch all bases in this overall project. But there is one thing that seems fairly clear to me, and that is that this effort will fail if it does not have credence amongst the academy. One part of the effort, since Goal 5.5 does say "college graduates," implies that college has something to do with what does or doesn't take place relative to critical thinking, problem solving and oral and written communication. We want to know the extent to which colleges and universities are leaning towards these concepts and skills currently. . . . And what we want to know, then, for example, is the relationship between those abilities as people perceive them and their relative importance to their campus and what their campus is currently attempting to do to teach them.

Granted, the criteria derived from the delphi process may not reflect the workplace. Granted, the criteria may not reflect citizenship development concerns. Granted it may not hold credence with legislative committees for the government nor with employers. But, also it needs to hold credence with the institutions in order for there to be participation of students and participation amongst faculty in a national assessment. Otherwise, you're not going to get their cooperation.

We're very concerned that this process include faculty who are presently teaching in colleges and universities across the U.S. The kinds of constructs and indicators and measures and domains and various [results you reach here will] be submitted to faculty who are interested in the four areas identified by the national goals. They will identify the extent to which they feel the colleges and universities are working or can work at attempting to address those areas.

. . . Thus we will ask them to basically evaluate the importance, or the concordance I guess is a better word, of the criteria discussed in each of the groups here with what they are trying to do in their individual institutions.

Let's take this particular Critical Thinking (CT) group as an example. Out of the 400 there will be people who will be identified as being involved in or concerned with CT. And those faculty will then become the participants in this Delphi process. . . So, it's not their role to evaluate John Chaffee's work, so much as it is the appropriateness or the extent to which John's work will explain the efforts on their campus to teach critically thinking overtly and explicitly or implicitly and embedded within the existing curriculum.

Others Involved in the Delphi Survey

So, when we go out to the 600 or so individuals, we want to get responses that are meaningful. The product that then will be reported incorporates both: your thinking and that of practitioners in the teaching fields covered by the Goals . . . After we've achieved a draft instrument that is meaningful in terms of communicating with you all, then we will have focus groups of individuals who work at colleges and universities, take a look at the materials and ask does this instrument make sense? If it landed on your desk would we need to explain what a "such and such" is?

This project is not a survey of major employers or policymakers in the U.S. So, the employers will be employers that are identified by an institution as a major employer of graduates of their institution. And the government officials will be those identified by the institutions as those individuals they work with most closely. And the institutions will be institutions that represent a national representative sample of students, not of institutions, because the Goals speak to student learning.

It will go to 600 individuals, 400 of which are faculty, and 200 of which are people associated with higher education, organizations, agencies, states and employers, 100 of each. . . . And then we see a condensation of the work of the four groups. The groups will be asked to serve in an advisory capacity to the development of the survey instrument. Then, the 600 individuals will be asked to engage in a delphi process.

How the Center Will Work With This Conference Product

What we propose to do is to take the revised papers from those authors who've expressed their wish to revise them. We're uncertain exactly how to incorporate the information of the reviews of just one paper, so the authors might like to incorporate those comments in their revised manuscripts. And then we will take the work of these groups, and provide it to some 600 faculty, employers, and individuals responsible in the government for higher education.

There should be two packages of materials. What the packages look like we know somewhat, but some of it is dependent upon [what we see from here]. The first package will contain the papers in their revised form, and either the reviews or the revised paper with a note saying these were papers . . . So we're not sure beyond that level of specificity what that package will be, but at least it will be the total revised papers.

Your question is—take this group for example—What does the condensation look like? [Q: Yes, who does that?] We do that, in consultation with you. Jeff said he is going to provide us with the notes, the recordings, and his representation of what occurred, or his notes and Mary Ellen's . . . We will then ask participants of this conference to join advisory groups to comment upon the development of survey instruments. Following that advisory process will come the focus groups to insure the clarity of the instruments to those likely to complete them. Finally, we conduct the two rounds of delphi and report back to NCES.

Thus, we will try to prepare a representation of what has occurred that [you] people can respond to in terms of—for example, we've had in this group four frameworks to look at: One question, I think that would seem relevant, is which of these four frames approaches what we might envision as CT. It strikes me for example that [to Diane Halpern] your dispositions from your paper approximate some of the things that are aspired to in general education courses. So one point of evaluation might be those four frames. [Halpern, Chaffee, Mumford, Perkins] . . . Another point of evaluation might be the specific skills or abilities that are identified, for example, in John's [Daly] outline.

What we'd like to do is, after we've assembled and prepared a draft set of materials, that is a compilation of the work that is done by the four groups here, [is] to share that with the authors and reviewers to get their feedback.

What we want to do is go out with something that people feel accurately represents the discussion that occurred here, and accurately reflects the thinking of the authors who were in essence the catalysts of the discussion. So that when we go out to the 600 or so individuals we get responses that are meaningful.

[NOTE: The 1993 Penn State Delphi Project focused upon critical thinking, writing, and speaking and listening skills. A 1994 Delphi Project added reading and problem solving skills. The 1993 findings will be published by NCES in a separate document in 1994 and the 1994 findings in 1995.

PROBLEM SOLVING AND CITICAL THINKING, GROUP TWO - AUTHORS PERKINS, TISHMAN, AND JAY

PARTICIPANTS:

Facilitator:

Eugene Owen
OERI, NCES

Authors:

David Perkins and Shari Tishman
Harvard University, Project Zero

Eileen Jay
Consultant, Arlington, Virginia

Reviewers:

Magda Colberg
Office of Personnel Management, Research and Applications Division

Peter Facione
Santa Clara University, Dean of Arts and Sciences

Suzanne Morse
Kettering Foundation

Participants:

John Chaffee
CUNY Laguardia Community College, Creative and Critical Thinking Studies

Dennis Jones
National Center for Higher Education Management Systems

Jean McDonald
National Governors' Association

Bob Morrison
Navy Personnel Research and Development Center

John Muffo
Virginia Tech, Program Review and Outcome Assessment

Susan Nummedal
California State University, Long Beach, Dept. of Psychology

Ernest Savoie
Ford Motor Company, Employee Development

Others:

Chris Carr
Recorder, University of Maryland

Robert Burton
OERI, Naitonal Center for Education Statistics

Judith Seigel
OERI, Office of Research

Sheida White
OERI, NCES

PROCEEDINGS:

Introduction

Of the two papers commissioned by NCES on the critical thinking component of Goal 5, Objective 5, the one prepared by Perkins, Jay and Tishman (hereafter Perkins et al.) led the workshop into a discussion about how to emphasize salient points of assessment, many of which were raised in the original paper. Reviews by Mumford, Tenopyr, Colberg and Morse also addressed the Perkins et al. paper, and both Paul and Facione each provided their reviews in separate documents which addressed both the Perkins et al. paper and the other commissioned paper on critical thinking by Diane Halpern. Of these reviewers, Paul, Facione and Colberg were members of the working group. Additional group members were Larry Sutter, Dennis Jones, Jean McDonald, Bob Morrison, John Muffo, Susan Nummedal, Ernest Savoie, Robert Burton, Sheida White, and Judy Seigel. The group discussions were facilitated by Eugene Owen.

The group's discussions embraced the paper, its exposition by the authors, all of the reviews, and the elaborations provided by those reviewers who were actually in the group. Another element also drove the debate: the news that the National Center for Postsecondary Teaching, Learning and Assessment at Penn State, under the directorship of James Ratcliff, was to use the results of the conference to develop a questionnaire which would—through a modified Delphi process—be the basis of a survey to be sent to some 600 people in the national community. Finally, the group was able to structure their discussion with the icon suggested by the organizers. These are the components—these participants and this general body of information—which led Group 2 to produce several discrete "products."

- ○ A rich conceptualization of what critical thinking is.
- ○ A list of competencies necessary for effective critical thinking.
- ○ A list of critical thinking skills.
- ○ A list of Orienting Points for Penn State as well as a set of General Principles that any others in the process should be mindful of.

Each of these will be described in turn, and then a synthesis will highlight those aspects of the discussions that seemed to group members the most critical and which seemed to indicate underlying themes and connections.

The Nature of Critical Thinking

The Group 2 discussion provided yet another staging of the ongoing debate over the theoretical constructs of the critical thinking movement. The paper by Perkins et al. captured an approach which emphasizes the process over the product of thinking, as well as the crucial nature of the context in which any thinking is done or evaluated. This viewpoint emphasizes the inherently psychological aspects of any problem solving situation. By contrast, Paul (whose work and theory were discussed in Greenwood, 1992)—a self-described "philosopher by trade" —believes that the underlying foundation of any successful taxonomy of analysis must be epistomologically based; that is, they should have a systematicity and internal logic that make them a tool with which problems in any context can be approached. The review by Paul et al. exemplifies how this might be accomplished, and how they believe the Perkins et al. view falls short. Facione's background as the author of the Delphi report brings yet another view to the table, which he believes is fundamentally different in that the empirical predominates over the conceptual (whether Paul's more philosophically based, logical approach or the emphasis on process and context by Perkins et al.). Like Facione, the other reviewers provided observations which set the stage for Group 2 discussion: an effort to move beyond the debate over systems to some sort of underlying core, as well as recommendations of how to proceed with assessment.

What are the hallmarks of critical thinking? The epitome of critical thinking is judgment, a quality evident only through the process of establishing a non-trivial question and then reasoning through to an answer. This procedure is not simplistic and often depends on the process of meta-questioning to clarify the issue and to identify point(s) of view and frame(s) of reference involved. The successful effort to frame a difficult question sometimes leads to a straightforward answer. Alternatively, the issue may be multidimensional and require the reasoner to develop a more complex frame and approach. But whatever the process by which the question/problem is approached and framed, the critical thinker is consciously seeking a solution—is purposeful—and thus regulates her or his own progress through the problem. Such awareness includes a sensitivity to: the interpretations made of that information which is selected; the concepts and assumptions employed to generate solutions; and the implications and consequences of these findings. Thus did Group 2 define what might be called the structural elements of critical thinking.

Yet, critical thinking tends to be an undeniably social act. Even when framing or addressing a problem alone, good critical thinkers embrace the social context of the problem in order to reason through to the best solution. More often, critical thinking requires direct interaction with others, and thus the issue of self-regulation moves to a more complex level where the social dimensions and implications of the answer feed back directly into the process undertaken in discovering it. A good example raised in Group 2 was the continuous awareness group members expressed about the social implications of their assessment recommendations.

Competencies Necessary for Effective Critical Thinking

The group agreed that a core of basic cognitive competencies underlies all performances: Interpretation, Analysis, Evaluation, Exploration, Self-regulation and—possibly the most global—Inference. Most often, the performances are directed to problem solving and employ basic ways to approach a problem: Design, Synthesis, Explanation, Decision-making, Justification and Explanation. These stages are not mutually exclusive and may overlap or repeat during the process. Collaboration was emphasized as continuous through most of the stages, by enlisting others in the process as well as listening to all other relevant actors.

Echoing the above characteristics of the nature of critical thinking are some often-found things successful critical thinkers do: Think in a broad and adventurous way; sustain intellectual curiosity; continuously try to clarify by seeking knowledge and understanding; seek truth and evidence with intellectual courage, integrity, and responsibility; think in ways that are planful and strategic; maintain the rigor to be intellectually careful; and remain continuously aware, thinking about the process of thinking, i.e., metacognition.

Closely related are the dispositions critical thinkers bring to the act, the attitudes with which all of these other activities can be approached. Good critical thinkers are Inquisitive, Open-minded, Systematic, Analytic, and always Seek the Truth through Objectivity and by Reasoning. They are also socially attuned to Factoring in the critical thinking of others, and to Using collaborative skills while demonstrating Intellectual Empathy. Finally, they have what might be called the elements of Critical thinking self-confidence: Courage, Cognitive/epistemic maturity, Intellectual integrity and Intellectual responsibility.

The Critical Thinking Skills

In the real world, the Critical thinking competencies are manifest when people are solving problems, an act which often involves the thinker in a three-step process: Querying a body of information or evidence; Postulating hypotheses and alternatives from this data, and Drawing warranted inferences.

The first crucial step is to seek the question to be solved, in essence to recognize and frame the problem. While this is a threshold issue, it can recur throughout the process, requiring a number of skills such as Resolving competing values and a willingness to reconceptualize, to redefine the problem along the way. Generally, resources must be marshaled and assessed, and goals must be set and prioritized. This requires continual awareness of the feasibility and the long-range implications of the solution being approached. Each item in the string of reasoning must be evaluated critically, and integrated within and connected to the overall framework. These results are obtained by continually communicating and coordinating with others involved, and getting them to "buy in" as the process unfolds. Persistence is crucial throughout.

What are some of the crucial, distinguishable steps taken during such problem solving exercises? Assessing and clarifying the purpose and then identifying and critiquing the question to be addressed, or the problem to be solved. At some early stage it is important to identify and to critique both the assumptions being made, and the concepts involved. Gathering information, and judging its relevance, are fundamental skills, consequent on which reasonable inferences and conclusions may be tentatively drawn. The next important step is to Identify and assess the implications and consequences of these, challenging, questioning and giving relevant reasons for them. This may take the form of Identifying and reasoning hypothetically within alternative points of view and frames of reference.

It is one thing to engage alternative points of view, but successful critical thinking in a social world requires a number of skills. Much of this behavior occurs not only in a group context but within a group of people that could be called a team, with an ostensible common goal. Establishing and configuring such a team, feeling and seeming to belong to it, informing it with a sense of purpose, resolving internal conflicts, and building consensus within the group are often essential aspects of effective critical thinking in groups.

Three particular elements were stressed in the discussion of the most important skills. On the one hand, the importance of logic as a tool. This includes recognition of and remaining compatible with the logical schema that is relevant to the problem at hand, as well as applying throughout the entire analysis the essential glue of logical reasoning. On the other hand, critical thinkers manifest creativity, a quality as difficult to describe as to teach. Most good critical thinkers embody it, however, employing techniques like systems thinking and brainstorming, for example. Both creativity and disciplined logic are important qualities which come out in the wash of self-examination. Only by a continuous process of individual and group awareness can the ongoing and crucial self-corrections be recognized and accomplished.

Products for Penn State

Several of the group members have produced scholarly work that anchored the discussion of these various skills and abilities, though the group as a whole became a keen lens through which to review how these elements may overlap and fit together. In working sessions, others in Group 2 considered the next step of the process, where the Center for Postsecondary Learning and Development was to take the results of the conference to a larger audience through a survey yet to be developed. Since so many controversial issues of assessment involve perception and emphasis, this subgroup developed two documents intended to orient analysts at Penn State and any others about some of the crucial elements the group foresaw in the future of the assessment. One was synthesized by John Muffo, the other provided in the form of a list of Orienting Points for Penn State.

The list is available in its entirety, but some crucial points that apply to the overall results of the NCES conference were especially relevant. The group found the use of the icon to develop lists of skills to be a useful strategy, but wanted to be certain that it not be reified. There will necessarily be redundancies in any group's list, and between groups, and care needs to be taken, in trying to pare them down, to examine what is meant by the various entries—especially since many of the skills intersect in several categories. There is a great deal of literature on these terminology overlaps and distinctions which should be consulted, including SCANS, other governmental classifications, and the material developed in the earlier conference. And perhaps the group's most strongly felt caveat: the expertise to make and appreciate such niceties resides in the experts, a sample of whom were at the conference; the Penn State step or any important subsequent step, should build in essential consultation with this community.

Orienting Points for Penn State

Group 2 worked to develop an explicit document to convey these ideas, which was presented at the end of the meeting to the assembled conference. Here follows the document as presented, with annotations by David Perkins:

> **These are orienting statements to accompany the taxonomy on the icon from critical thinking/problem solving Group 2. They are meant to provide a context for making the best use of the taxonomy in coordination with taxonomies from other sources. The order of statements is arbitrary.**

We felt it was important to frame the expanded taxonomy by a number of ideas or statements that would help to orient Penn State and others who will say it, as to how to take it, how to understand it, and how it might best be used. We also suggest that, in general, the products of all the groups might well be framed by some cautionary statements, interpretive statements. One of the reasons we think this is important is that, inevitably, the products of a group process like this where you only have a couple of days, are rather rough-shod, and still contain within them conflicts, ambiguities, and the like.

So, in that spirit, I'd like to introduce to the entire group here the statements we thought were important in our context, and invite others, perhaps, to do likewise with the products of their own groups. Let me make it clear that these orienting statements are not mine, but rather the results of our group's process. I happened to be the person in the group with a laptop, so I got stuck typing them up, composing the detailed language. I'm not sure it's worth the group's time to go through these items one by one. Let me see if I can telegraph the gist of the message, and perhaps underscore a couple of items that are a little different from the rest.

> **The taxonomy includes a number of seeming redundancies—categories that seem to mean almost the same thing. They appear because some of us saw important contrasts in ideas that sound somewhat the same. WE agreed because of time not to try to reconcile these terms.**

The gist of the message of a number of these items is that, quite simply, within our group products, you will find overlaps, ambiguities, categories that seem to apply to other categories, terms that may not be clear, seeming redundancies. The basic message for all of those is that we didn't have enough time, and therefore we chose not to try to refine out some of the overlaps, to leave things a little rough. We chose not to clarify terms on these pages, but much of the clarification can be found, not only in the paper contributed to Group 2 but in other papers contributed to this conference and the previous conference. All of this has a cautionary character. All of this says, basically, "Bear with us, this was as far as we managed to get."

> **This taxonomy is the result of a group process where we choose in many cases to pool ideas rather than resolve disagreements. Nonetheless, we did make some effort to sort categories in an orderly fashion and conceptualize the whole in a reasonable way. This should not be viewed as just a pile of ideas.**

But, the added message is that this was a serious effort. We did try to achieve some coherence. We did try to place things in at least rough relationship to one another, so if things seem a little higgledy-piggledy here and there, please take seriously the general gist and structure. That accounts for a good number of the items.

> **We call your attention to our attempt to provide a schema that made reasoning and categories of problems a central organizing concept to make greater sense of the whole. This schema can be found in the notes from our group and begins with the words "reasoning about some question at issue."**
>
> **Many of the categories in this taxonomy have relevance to other categories. This should not be a worry. This is the way thinking is. For instance, seeking clarity has relevance to justification and justification has relevance to seeking clarity.**
>
> **Some categories in the same column are much broader than others, and some categories are elaborated into subcategories whereas others are not. This is by and large an accident of what we had time to attend to.**

Let me underscore, just scanning here, a couple of others.

> **We see great generality in these categories. They are relevant across most or all disciplines and contexts.**

We felt that it was important to underscore that, in our group at least, we don't see these categories as simply being reduced to discipline-specific concepts: they have generality.

> **Some of the terms in this taxonomy may not be clear. Many of them are clarified in the contributed papers and reviews from this and the previous conference.**
>
> **We urge that this taxonomy be examined in relation to SCANS and other national taxonomies from the government, in addition to the taxonomies previously proposed.**
>
> **We want richness and depth in the interpretation of this scheme.**

And I'm sure the same would apply to any of the schemes produced.

> **Simplification is desirable, but not at the cost of oversimplification. Thoroughness is important.**

There was some trepidation in the group about the fate as we hand these pieces of paper on, of the concepts and distinctions. And maybe one more, if we're underscoring.

> **The further process should be attentive to workplace, citizenship, and college constituencies, not neglecting any!**

We felt it was important to underscore that.

> **Dispositions should play a prominent role alongside abilities. According to the initial paper for this group, dispositions and abilities were paired one-to-one. Whether this was a good idea, however, was not resolved.**
>
> **Intellectual standards should be explicit and prominent in the final product. Intellectual standards should not just take the form of minimal standards.**
>
> **People with expertise in critical thinking and other relevant competencies should be involved in every further step, in particular in the Penn State process, to be balanced with strong representation from workplace and citizenship perspectives.**
>
> **Several taxonomies of critical thinking and problem solving have been suggested at various stages in this process. They are few in number. Although they have their divergences, they agree in many respects. We suggest that cognizance be taken of all of them.**

Let's be clear about that. If 30 taxonomies had been presented along the way, it would be unreasonable to suggest that those that follow after us try to sort them all together. They might just go with the latest output of this conference. But that isn't the case: there aren't that many on the table, and therefore it's entirely reasonable to expect the next stage of the process to look it all over and see what sense they make of it.

I don't know whether anybody in Group 2 would like to add anything to that overview of some important orienting remarks, but again I suggest that in general, perhaps the all of [the various groups'] products that go forward should have orienting remarks so that they will be taken in their best light, their flaws viewed gently."

General Principles to Penn State

Similarly, Group 2's larger list of principles was presented to the assembled conference at the final session by John Muffo of the Virginia Polytechnic and State University. Here follows the document's major points in boldface, and comments made in the final open session to elucidate:

JOHN MUFFO: We were in Group 2, which subdivided into a minimum of 2 subgroups. There were some points that the subgroup thought were important, and I'd like to just briefly run through these. There was a concern about the next steps, about what goes to Penn State, and what advice we can give Penn State. Not that our group wants to run the process for

Penn State, but if we could tell them something, here are the things we'd like to tell Penn State.

GENERAL PRINCIPLES TO PENN STATE REGARDING HIGHER ORDER THINKING SKILLS IN WORK AND CITIZENSHIP

First of all, the whole issue of terminology. We thought critical thinking might be good for us, but there do seem to be some negative connotations to the word *critical* in the public domain, and so we like the terms "Higher Order Thinking Skills in Work and Citizenship" better, especially the "higher order thinking skills" part. Also the emphasis on work and citizenship, add them to the title to focus attention on the purpose of doing assessment.

Basically, we broke these down into two areas, one being communications, the other being specifications.

COMMUNICATIONS

*** Emphasize the process over content knowledge.**

> **The focus of any assessment should be on the process of exercising higher order thinking skills as opposed to the inculcation of disciplinary information.**

I know a lot of the disciplines are dealing with this. My own College of Veterinary Medicine is having a real struggle with this issue, and they're going more for process vs. content knowledge, but I think part of what we should be trying to communicate to the broader public is that this is an emphasis on process rather than content.

*** The overall assessment process is summative, though with a formative intent.**

> **It is recognized that many of the driving forces behind assessment, particularly this assessment process, are due to a desire on the part of policymakers in the public arena for a summative/instrument process to evaluate higher order thinking skills.**
>
> **Nevertheless, a well-constructed measurement device should provide information that can be utilized by institutions of higher education and faculty to make improvements in the curriculum. The potential for self-improvement should not be overshadowed by concerns over the summative origins.**

We had near-consensus on [the notion] that, essentially, this is a summative process. Whether or not we would like it to be. But, instead of ignoring that, or pretending that's not true, I think we should say that. We should say it, and these are my words, but then let's go ahead and get some formative assistance from this process. In other words, if we do have a product that says "our score is 3.2 and the score of our sister institution is 3.3, how do we use that information in some way to improve our institution, or improve the learning process within our environment.

*** Continue the process with a follow-up conference.**

> **A concern was expressed as to how the larger audience of higher education, employers and public policymakers will become aware of the project. The act of contacting a wide variety of individuals through a Delphi technique as a survey will itself raise the visibility of the project, create questions, and create a demand for resulting data.**
>
> **A positive way to build support for the project—while ferreting out and addressing possible resistance—would be to hold a conference which highlighted the results of the first survey and which would bring together the key interested parties, such as representatives of the higher education, employer, and public policy communities.**

Our group expressed a near-consensus that this concept needs to be sold. The world needs to know about it. The very act of going out with a quasi-survey, of going out beyond this immediate group, will raise expectations. People will be calling and saying "What's going on here? What's happening?" One of the suggestions we had was that, when there are some preliminary results available, to call together a group of the key interested parties. Meaning not only institutions of higher education but employers, public policy people, governmental folks, to just kick around the issues amongst themselves. This group seems to be predominantly higher education people, and we may be preaching to the saved, here. So the idea was that this would help elevate the whole process to a higher level of publicity, to assist NCES in, first of all, getting support, and second, finding out where the opposition is, so to speak. And seeing if there is a near-consensus. The more people we can get involved at this state, particularly of the customers out there, the better. At some point, we have to quit treating this as a secret.

SPECIFICATIONS

*** Make it a marketing device . . .**

> **Part of the role of a survey would be to begin the process of informing the future users of a resulting measurement instrument of the existence of the project. In addition, gathering the opinions and needs of these potential users will help to better determine how to market the final product back to them. This act of information gathering is a type of market research.**

. . . which will be explained to the "clients" or "customers."

> **The term "client" is preferred over customer in that it conveys the idea of both recipient and supplier of information moving together, while "customer" implies movement in one direction only—from the supplier toward the recipient.**

Basically, we saw the entire process as something of a marketing device. In going through a survey, for instance, we would have to be sure that we're explaining this to the clients or the customers. We use the term clients rather than customers, because we see that there's a gap between the higher education community and the rest of the world out there, and the client image much more suggests moving toward each other rather than either of the parties having to change dramatically. [Again, those of you on the subcommittee, I'm probably forgetting some important points here.]

*** Make it simple and parsimonious; translate as much as possible into common terms.**

> **The concern is that any terminology used in the survey or public discourse about the assessment be expressed in terms easily understandable by the nonacademic layman. The basis for all work must remain conceptually sound and defensible, and thus it presents a real challenge to convey complicated matters lucidly and understandably.**

There's a real struggle there, and it's something we have to work on very hard. But we need to try as much as we can to put these very complicated concepts into layman's terms, and maybe give them some explanations so that the layman understands what we intend when we say something.

*** There needs to be, to the extent possible, a focus on interdisciplinary work, with particular attention paid to the proverbial stepchild of the curriculum, general education.**

*** The kinds of problems involved must be clearly identified, along with levels of achievement.**

> **One of the crucial aspects examined by an assessment should be problem identification, clarification, and selection. How one goes about this process, and the range of types of problems addressed, can be critically important. The outcomes of the assessment will be more useful if expressed in achievement levels.**

We've been talking about what kinds of problems, and there's the problem of identification in this area is particularly important. If we're going to say, "This is a minimal standard, this is a subminimal standard, and this is someone who is above the above-minimal standard," then, while I hate to say grading scales, we will have built this is as a more useful kind of information, in some regards, than saying simply, "Yes, you pass," and "No, you don't pass."

*** The kinds of evidence produced must be credible and useful to the clients, showing what particular skills are demonstrated and how.**

> **In order to convince the clients and the public of the usefulness and applicability of the eventual instrument, the process should ensure that the target audience identifies the skills they believe important. The assessment results must include evidence of the acquisition and demonstration of these skills.**

I think here we're just arguing that any type of test or testing procedure should demonstrate skills as well as knowledge. In other words, being able to do something, as well as knowing something. So that's the what, but there's also the how: How do you demonstrate that? One might be tested on a certain skill, but then the more difficult issue arises: how do you really test that skill? So we want to know if it's present, but if it is present, how do we find out?

PETER FACIONE: John, on the kinds of evidence, I think the point there was, "What sorts of evidence would be persuasive to the business community and the policymakers? Not just what sort of evidence would be persuasive to psychometricians and academics. What's persuasive ultimately to the clients, who are the policymakers and the workplace employers.

JOHN MUFFO: Okay, that's the issue of face validity that we discussed earlier. Okay, thank you.

*** The sample for this next step should be ecumenical and representative.**

> **The survey should include the universe of baccalaureate granting institutions of higher education as well as the employer community, young alumni, community leaders and policymakers.**

If we do go out with a survey or quasi-survey, whatever we're calling it, we thought it was important to go to a wide range of academics and institutions; not just major research universities and not just public institutions. We thought it was important to involve employers of the various kinds: private, governmental, non-profit, and so on. [The next point was more or less mine, I guess.] I think it's important that we ask young alumni some of these questions, because we're going to get different kinds of answers (I'm reflecting on my own experience, now) than when we're asking employers or other groups. Because these people are the ones who have found themselves lacking or not lacking in certain areas. I think their experiences would help us to understand, particularly with the entry level, the skills and knowledge levels that are necessary to succeed. And then the community leaders and policymakers; essentially these are the folks that are driving us to this process in the first place. Community leaders and policymakers represent those groups behind the National Goals Panel and its resulting activities, so their opinions are critically important.

*** Several pilot projects should be mounted.**

> **Even the best-intentioned and best-planned projects can lead to disappointment. When pilot projects can feasibly be developed, doing so with multiple versions will ensure that the process is not reliant solely on a single model or approach.**

We felt that multiple pilot projects were important. If anything goes forward, a single design would probably not be best. Rather the group thought it wiser to try several different approaches to see what works best. There may be some abject failures, and that's okay, as long as we don't rely on just one design. There are also advantages to being able to choose from among several options.

*** The instrument and the process should be founded on a multiple measure approach.**

> **When an assessment device or approach is developed, a variety of measurement methods, measures, and types of evidence are strongly recommended. A wide variety of higher order thinking skills are measurable by various different means. To truly measure a range of skills, a range of measurement techniques must be developed.**

We're looking for multiple measures and types of evidence and methods and so on. Again, the whole issue is that a single measure is not going to capture more than just a small proportion of all this.

***A clear set of priorities must be established and communicated.**

> **Since a large number of higher order thinking skills is likely, forward momentum in the process could be hindered by a lack of focus, both in those responding to the survey and those developing the assessment instrument and techniques. Any information gathering process should have respondents establish their priorities for the skills in question, in order to refocus emphasis on those matters the clients decide are most important.**

There are so many things that can be measured here. I mean our group wasn't able to come even near an exhaustive list in two days. It's going to require testing and evaluative procedures in order to set some priorities about what are the most critical and important issues, skills, levels of knowledge, and so on.

Overriding Themes

The foregoing two documents grew out of Group 2's general deliberations on the crucial issues involved in developing the assessment. Perception was a central concern. A first step was to realize that the move to the larger audience will involve people who are not necessarily aware of the body of thought and research underlying the critical thinking movement. Thus, the focus of the assessment might better be phrased "higher order thinking." This would begin to accomplish what the group believes was an important strategy: making the message simple and parsimonious, finding lay terms to displace technical ones (where possible) to communicate more effectively. **If abilities demonstrably tied to work and citizenship are to continue to drive the assessment, then emphasize these by adding them to the title of the project: Higher Order Thinking Skills in Work and Citizenship. Offsetting the value of this wider and presumably more welcoming terminology was the possibility that it symbolized the beginnings of a process where the substantial work in the field to develop and describe a rich conceptualization of critical thinking could be jeopardized.**

Another central strategy was to keep foremost the ultimate context in which the assessment will itself be assessed, and to be sensitive to the need to market the device effectively; to perceive all of the affected audiences—educators, policy people, and the public to be tested—as clients or customers; and to treat them with the associated sensitivities that prevail in the marketplace. Among other things, this means to discover the actual levels of achievement and the kinds of problems these "clients" want and need. While the survey is a good beginning, the resultant exposure and higher visibility of the project will generate questions, curiosity, possibly concern, and a focus on the survey results. The approach to consensus building already evident at NCES should be continued, and a large public conference should be held to bring together the interested parties who were surveyed from the education, employer, and public policy communities. This is but one way to achieve what the group felt to be a crucial element of the development process: the need to keep it open and dialogical.

Implicit in this ecumenical approach is the recognition that multiple measures and differing types of evidence will be needed. And with any such schema, and because so many different target skills are likely to be identified, it becomes crucial to establish priorities. It seems hard to imagine taking this step at the present stage of knowledge and feedback from these various constituencies—even after the proposed conference process—and so pilot testing would seem an irreducible element of any assessment development. Sample audiences that might be tested are: a broad sample of academic institutions; all types of employers, including private, government, and non-profit; young alumni closest in time to both the experience of education and the struggle to adapt to the world of work and citizenship after graduation; and, continuously, with policy leaders at all levels, including the local community, as well as on Capitol Hill.

Though this process reaches out to what the group considers the essential parties involved, at the same time it carries the paradoxical danger that the price of such political diversity will be the loss of aspects crucial to its success. A number of pieces of the puzzle are inherently at cross purposes. While not necessarily offering "how-to" solutions to these conflicts, the group drew the battle lines and (in some cases) expressed their value preferences.

- While the ostensible goal of the workshop was to produce a list of skills that might form the target of the assessment, the Perkins et al. paper eloquently stated how such "end products" could mislead developers to lose sight of the crucial underlying process involved in all higher order thinking.

- However, the desire to make the process politically successful and create a broad sense of "accountability" tends toward discrete, testable elements. Any such heavily political and broad-based undertaking must have face validity, and employers will buy into this the more such elements are perceived to be related to workplace success.

- Also possibly at odds with discrete testable elements driven by workplace demands is the present educational infrastructure. No national assessment will fly without the active support of faculty and institutions, and a law of human nature probably suggests that dramatic change here must be very carefully wrought.

- Perhaps the most difficult such question of approach is whether to structure the assessment to provide a broad-based description of the state of higher order thinking, at the expense of a more intricate tool that would feed back to, and inform, the school on how it might refine its instruction.

- Another infrastructural dilemma is the departmental nature of most institutions. Interdisciplinary connections and so-called general education efforts often come in second and third to the perpetuation of discipline-based knowledge. For better or worse, this is how most evaluations have been structured, and those that go into so-called intelligence testing are under siege as more indicative of a student's background than thinking powers.

Richard Paul provided a model which the group felt a useful place to begin to "think your way" out of the inherent paradox of political pressure versus philosophical integrity. The underlying premise was to develop a commonsense approach which could be adapted to appeal to the various actors in the process, and thereby avoid the resistance that can be expected when people and/or institutions perceive a direct threat.

The first level of the model postulates that three fundamentally different types of problems exist:

Civic	Personal	Professional
Broad social, moral and political issues and problems	Everyday personal problems and questions	Technical problems and issues

Next, that the fundamental definition of critical thinking as a process (as earlier described) can be distinctly applied in each area, using approaches selected from a continuum:

(along a continuum)

Broad or general structures, employing broad dispositions, abilities and standards:	Highly specific and technical structures:
These global skills tend to arise from holistic abilities that can be applied to reasoning in many diverse fields or contexts.	These skills may be inherent to the nature and context of the problem to be solved.

The perceived value of such a model is that it permits the assessment to be sold to a number of different constituencies—perhaps most importantly to university faculty—with the understanding that it will accommodate present orientations and emphases. Developers can then mediate how the actual test sections might be constructed relative to the elements portrayed in the chart, without sacrificing the integrity of the thinking process itself.

- A certain mastery of reasoning and logic appear crucial to all higher order thinking. However, current educational practices have neglected this irreducible element, and hence test developers must balance the need for remediation with a test that would shoot too low.

- Yet there is an everpresent danger of creating an overly specialized or esoteric instrument that neither serve the political need, nor be immune to charges of cultural bias. Thus the assessment must strive for simplification, though not at the expense of what Paul and others have termed "a rich conceptualization of critical thinking."

- The more philosophically oriented researchers emphasize the central role of intellectual standards. They believe that reasoning should be assessed for its clarity, coherence, consistency, specificity, precision and accuracy, objectivity, significance, depth, breadth, adequacy to the task, fairness; its self-awareness and imagination; and its logical discipline (which includes the logic of its connectives as well as the overall schema employed). Specifying these standards in assessment structures will probably be a controversial and technically challenging task.

- While all agreed that effective higher order thinking cannot be separated from the dispositions they identified as part of the process, they could not agree on how to assess these dispositions, and whether the dispositions could be matched and realized through the demonstration of particular skills.

SPEAKING AND LISTENING SKILLS, GROUP THREE - AUTHOR - DALY

PARTICIPANTS:

Facilitator:

Barbara Lieb
OERI, Programs for the Improvement of Practice

Author:

John Daly
University of Texas, Austin, Dept. of Speech Communications

Reviewers:

Gustav Friedrich
University of Oklahoma, Dept. of Communication

Don Lumden
Kean College of New Jersey, Dept. of Communications and Theatre

Andrew Wolvin
University of Maryland, Dept. of Speech Communications

Participants:

Ed Morante
College of the Desert, Dean of the School of Education Resources,
Research, and Technology

Gregory Munro
University of Montana School of Law, Professional Skills Development

Barbara Pagnotti
Kidder, Peabody, Investment Services

Rebecca Rubin
Kent State University, School of Communication Studies

Karl Schilling
American Association for Higher Education

Orlando Taylor
Howard University, Dean of the School of Communications

Joseph Zeidner
George Washington University, Dept. of Administrative Sciences

Others:

Virginia O'Keefe
Recorder, George Mason University

Walter Chiavacci
Horace Mann Learning Center, U.S. Department of Education

PROCEEDINGS:

Introduction

Group 3 convened to discuss assessment of speaking and listening in the light of author John Daly's paper and the attendant reviews by group members Don Lumsden, Andrew Wolvin and Gustav Friedrich. Barbara Lieb from OERI was the facilitator for the group. While the acts of speaking and listening would seem to be the oldest of human cultural activities, canonized over two millennia ago by the Greeks and Romans as the art of rhetoric, there is not as much consensus as one might suspect about the components of these skills, given the myriad purposes to which they are put and the contexts within which they arise. When the question on the table is how to undertake to evaluate them among—or teach them to—college students in America, yet another layer of consideration arises. As Daly pointed out, most students have not encountered a structured course in communication before arriving at college, and thus if the skills *are* present, they are most likely an epiphenomenon of the K-12 experience. Many in the group believed, however, that a more accurate snapshot of America would show many of the most basic communication skills to be underdeveloped, in the colleges as well as the workplace.

Another confounding factor the group appreciated was the cultural diversity that could be masking possession of many of these skills, as well as the fact that most seat-of-the-pants evaluations at the college level are primarily critical: complaints that students lack the apparent ability (skills) to engage in the learning process as envisioned by the teacher. A properly dynamic definition of effective communication, the group noted, would suggest that such failures, where they exist, cannot be limited to the "student" side of the relationship. Because Group 3 was stocked with a number of scholars who have been involved in teaching and studying the speaking and listening domains, many were familiar with the recent efforts in the field to analyze and organize the basic insights. Thus, rather than refine or ratify one or another of these existing catalogs, their more pragmatic view of the task before them was to provide something upon which a truly muscular assessment could be built. Rather than grapple with the intricacies of how to assess, they chose to construct a model of communication around which any assessment should be built.

The discussion below reflects the basics of this model. A number of phenomena inhere to the basic act of communication, framing any discussion of effectiveness, instruction, or assessment. Once a communication act is undertaken, it often falls into one of two categories that, if not distinct, are usually driven by distinguishable purposes: to inform or to persuade. Clarifying what is similar and what different in these two basic situations led the group back to a general principle which overarches the entire domain. To wit, that all communication acts involve a relationship between and among speakers and listeners. Appreciating the subtleties and mastering the conduct of this relationship is the *sine qua non* of effective communication.

Communication, in General

Communication is the act of transmitting and receiving created messages. And the actors are on center stage, engaged in an indisputably *personal* encounter. Thus, any discussion of this domain will revolve around relationship. Situations differ, of course, but a number of elements affect the writing/speaking relationship that do not have the same impact or relevance in the other areas targeted by Goal 5. Most of these elements important to speaking and writing can be categorized in one of two major frames: either the function of the encounter, or the setting in which it occurs.

Two people meet—the most basic of human situations—and yet about to unfold is the most complex of encounters, one far beyond the ability of scientists or computers to predict, model or fully describe. It is immediate: the people involved are *there*, and cannot avoid the event which has begun at the moment of encounter. It proceeds to unfold in real time, which moves only forward: every step down this communication path is irretrievable, if not irreversible. As in a dance, one person responds to another, at the necessary moment. Not orchestrated by actual music, these moments are nonetheless timed precisely and orchestrated by intuition (which is fired by some obscure conjunction of social convention, the human brain and its inherent cognitive and emotional timing, and other factors too complex to pin down). But the dance is ongoing ("The game's afoot," said Sherlock Holmes), and so the timing of the encounter is nearly inexorable, and thus evaluation of your partner's "steps" must be limited to what you can accomplish within this time frame. Too much elapsed time will change the timing, the tune, the dance, the entire communication encounter. Thus, your involvement as a speaker or listener is cemented in time, simultaneous (no tape delay) and concurrent with that of the person(s) with whom you are communicating. The prospect of falling behind can produce an interesting emotional subtext, prompting people toward disclosure in order to "keep up their end;" or to anxiety, lest the dance not go smoothly. On the other hand, communication is interruptible. You can stop the dance, change the tune, by attempting to provoke, encourage, or negotiate an adjustment of the situation with your communicating partner.

Thus is spoken communication interactive. It has always been *between* people, hence the ancient roots in human culture of the dialogue, back to Socrates and further. It is people that do this, and their human, physiological equipment is the medium through which it is accomplished. Mouths and tongues and throats produce sounds (mostly phonemes), ears hear them, and eyes and other senses gloss the transmission and receipt of the message with a general awareness of (a "take" on) the humans at either end of the communication tether.

Such awareness makes the communication encounter inherently intensive. People seem to have a built-in tendency to discern or derive meaning—this could serve as a definition of the social brain. Regardless of how subtly or flagrantly, most transmitters of a message have that purpose foremost: to convey the message. To achieve that purpose, they must construct and reconstruct its delivery, specifically to meet the situation presented by a particular listener(s). And for the listener, whose purpose most often is avowedly to receive the message, but who in any case is irreversibly engaged in the act of communicating—for both speaker and listener—their communication is a co-creation.

How to approach the communicative act analytically? Group 3 developed several ways to frame the task. Three primary domains or dimensions of human activity are concurrently invoked when people are speaking together: the communicators are processing and understanding (more or less successfully) the cognitive content of the message; they are also behaving, call it performing, as they speak and listen; and they are feeling, however subtly: presenting affects, developing attitudes, displaying dispositions.

The group identified four categorical purposes that people seem to be about when they communicate. First, they are transferring information and knowledge to one another. Second, one asks Why? Often (but not always) they do so in order to influence, convince or persuade others, though a campaign of successful persuasion can be predicated on withholding information as well. Conversely, people as listeners put themselves into an evaluative mode, resisting persuasion, or allowing themselves to be influenced by a communication. A third thing that happens, inherently or explicitly, as people communicate is that they become involved: relationship is inherent, and the purpose of much communication is to develop, extend, enhance, define and to probe the nature of the larger relationship between the communicators. This third purpose often comes clothed in rituals, referring both to the content and the conduct of the communication. Finally, people seem to want to have fun; much communication is explicitly or implicitly entertaining, and often this purpose ascends to the fore. Obviously, these four purposes can, and often do, overlap and subserve one another. The communication domain is much more susceptible to a holistic, Venn diagram than to linear lists and categorical outlines.

Onto this broad picture, wondered the group, what criteria might be imposed to provide a platform for analysis? Should communicators be scored as to how *effective* they are, assuming the parameters and indicia of such effectiveness could be agreed upon and then measured? Probably so, but then exceptions and caveats immediately arise. What about *appropriateness*? Also an important, common, nearly universal component of good communication. But not quite: situations occasionally call for inappropriate behavior in order to influence (effectively, if negatively) someone *away* from something. Combine these two into a rough composite, however, and you probably have what the group thought could be a fairly muscular parameter of judgment.

Appropriateness is easier to define and approach through the back door. To avoid pushing people's *inappropriate* buttons, communicators need to know the social norms and conventions that apply to their situation. They need to know as well what their listeners consider relevant, and to be ready, willing, and able to adapt to these expectations. Effectiveness is not so simple to get a handle on. Few communicators are consistently effective when they break the first two commandments of clarity and efficiency. And a virtual prerequisite for good communication is that it have impact. If not merely fleeting, the experience then becomes memorable and has achieved effectiveness almost by definition.

How, then, would one begin to evaluate communication on the Effective/Appropriate scale? While these aspects of a communication act are vitally important and to some extent overarching, it may be difficult to impose them *a priori* to the more particular, practical, piecemeal, context-driven, outcomes-based framework of analysis necessary to capture a coherent picture of the complexity of a communicative act. Such a way of analyzing and evaluating communication will be sketched out in the next three sections on Informing, Persuading, and Relationships. Nonetheless, some general features of the Effective/Appropriate scale are nearly universal and should be emphasized.

Trying to generalize much further about communication soon runs aground, because goals (somewhat more specific and immediate than the purposes earlier described that sketch a meta-view of communication) lead one to the defining influence of context, and into the subjective realm of outcomes. Contexts are many: meetings, presentations, giving directions, explanations, instructions; people in the act of reporting, interviewing, sharing, answering, solving problems, explaining, conveying information, and variations on all of these to adapt to various public forums and electronic media. These many different contexts often do not assume primary relevance *per se*, but rather a relevance based in the context of the purpose (informing, persuading, etc.).

Can we generalize, finally, about a successful communicator? About the competent (effective, appropriate) communicators any college assessment would try to identify and describe? Such people seem adept at using communication to accomplish their own ends or goals, and comfortable in doing so. Rarely are they disorganized and confused, hence confusing. They are generally confident, it follows, and willing to take risks, unafraid of exposure, contradiction or criticism. They are good at engaging people and seem sensitive to the needs such other people have when communicating. Most people seem to be "built" for the appeal of stories, and good communicators often rely on this.

Finally, what do successful communicators accomplish? If communication is the act of transmitting or receiving a created message, successful communicators seem to embody a credo for how to do this. They perceive the human drama as a series of relationships, which people generate and maintain by communicating. The dialogues they develop serve to deliver and obtain information, to articulate and clarify questions and dilemmas, often to solve problems. During these dialogues, they describe, differentiate, argue, reason, often entertain, occasionally deceive. But the message as delivered and received is all, providing a metaphorical river on which civilization flows.

Informing

The most common of all communications, giving and receiving information, some might call the foundation of civilization; it is certainly its engine, for without it there could hardly be rationality, order and progress through science and technology. What are people doing when they transmit information, how do they do it, and what happens on the receiving end of such communications? (As the group intended, the basic criteria of effective/appropriate can be imposed where it seems pertinent.)

Good speakers seem to move, more or less in sequence, through four successive stages. However, because of the complexity, subtlety and inherently interactive nature of the communication act, any of these may loop to any of the others, as feedback from the listener and insights in the speaker warrant an alteration of the course, midstream as it were.

First, most informing acts manifest a goal, by definition. Person A has the purpose of transmitting a message to person B or to group C . . . Z. The decision to communicate a given message usually carries an explicit expectation of *when* this should or should not be done. (Bump into someone at the water cooler, and the time to relate a story you read in the paper about a toxic leak down the street is *now*.) If the message need be transmitted in pursuit of a larger goal, then the question of *whether* it should be done is moot. However, such larger goals always pose the decision to the informer as to which information transmissions will best (most effectively/appropriately) accomplish the larger goal. The many different contexts in which an information transmission can unfold may influence this decision significantly, or only modify the style and substance to suit. Often, a context will have built-in goals; a business meeting among colleagues about a company decision is an example where context and goal are virtually inseparable. One's audience and the context of the communication act also frame the goal in the sense that certain listener interactions may be built in or precluded, and goals must be set, possibly adjusted, accordingly.

Second, the informer must create the message. S/he knows the goal, context, and audience, and thus works within that set of constraints. Sometimes there is a choice of medium, but most often such variations entail how to present the message (third stage, below). S/he gathers, selects, and organizes the information needed for the content of the message. Decisions are also made about a balance between strictly verbal (words) and additional components of the message such as gestures, and other aspects of the message that will be directed to all of the senses of the listener. Just as there is a time to present the message, there is a time *for presenting* the message, often limited by the context and the nature of the audience. Creating a successful message requires the performer to understand all of these constraints, and to figure out the very best form the message should assume to be the most effective/appropriate. Clarifying the message is crucial, which usually requires a coherent presentation: a proper structure, foundation, transitions, usually "a point" (as in "Get to the . . ."), and appropriate summarizing and perhaps repetition. Again, the basics of storytelling apply to a great many communication situations. If possible, many informers will rehearse their performance: many-fold more marriage proposals have been rehearsed than delivered, and more delivered than accepted.

Third, the lights come up and the informer is *on*, presenting the message (see next paragraph). Fourth and finally, the informer assesses the results of the transmission. In formal presentations, this fourth stage can be fairly discrete, following the presentation of the message, as in Bill Clinton hearing and reading about reactions to his inaugural address. Most often, the encounter is more personal and interactive and thus the assessment occurs more or less continuously, throughout the presentation. Effective communicators realize the results of their transmission as quickly as possible, and often use feedback to actually reconstruct their message—or at least their presentation—in response to the listener(s), proof of the surpassing significance of relationship to speaking and listening communication. Managing this reconstruction process through questions and answers can be a complicated and strategic skill.

Texts have been written on presenting messages, and obviously the context/audience is such a significant element that generalizing here is risky. Group 3 did believe several abilities were to be found in most good communicators. With information transmission (almost as much so as in the next major section on persuasion) the bottom line is credibility. But the question really is how do people become perceived as credible, which properly directs the communicator's sensibility to h/is/er effect on the listener and the relationship that develops between them. Whether articulate, earthy or erudite, good communicators know how to use words and have "mastered the language" in the sense that the most effective words, phrases and gestures for the given situation seem at their command. Vocally, physically, and interpersonally, they are able to appeal to their listeners successfully. This is where the interactive dance of communication really happens. They must be in control, but not overbearing. They must be, or *appear to be*, willing to adapt in order to keep their listeners positively engaged: open and willing to receive the message.

And what about this listener, engaged in the receipt of information? the timing issue is paramount for listeners, because the message almost always comes but once, and thus their role as receiver is simultaneous and dynamic. The threshold question involves the desire or willingness to seek the information contained in the message. If one's goal as listener is more complex, then a suitable strategy for thinking or interacting must be adopted. Otherwise the good listener does just that, *listen*.

But how? Concentration, first and foremost. Next, one must have some appreciation of the speaker's many moves and choices (as outlined above) and be willing and able to follow and—where necessary—adapt to these. One element of effective listening involves distinguishing the various cues that are being employed to "influence" the receipt of the information. Such cues can be verbal, largely contained in the connotations of words and phrases; or they may be conveyed not on the cognitive level but the other two (felling and performing), through gesture, tone emphasis, and various nonverbal cues.

What about processing the content of the message? One does this privately, internally, and may or may not be able to engage the speaker in a dialogue to abet the process. Basically, the job is to store and be able to retrieve the information, to "get it" on the most literal level. To accomplish this (depending on the length and complexity of the message), listeners can effectively utilize internal summaries. Remembering the first commandment of information-giving effectiveness, listeners must develop and grant credibility. Since this is so crucial, and the human brain seems designed to discriminate, good listeners are continuously in the process of evaluating the inherent structure of the message for credibility: even where the message doesn't take the form of a structured or logical argument, good listeners are always evaluating sources of information and distinguishing facts from inferences, suppositions, innuendo, implications and specious conclusions.

Finally, listening to information can take an active role. In virtually all social situations, people are compelled to respond to one another. Listeners do this in line with their goal(s) in receiving the information. While good speakers are trying to control the environment and the dialogic situation, good listening at times calls for usurping this control. Framing and asking questions can be a crucial part of clarifying just how well one's internal representation of the message matches—becomes consensual—with what the deliverer of the information intended. How active listeners can and should be may well be limited (and often is) by what is appropriate. This anticipates the final section discussion about the relationship aspect of good communication. Ultimately, in most communication contexts, speaker and listener can agree to decide for themselves what is appropriate, regardless of the situation, social context, or convention.

Persuading

To persuade is to focus on the results of the relationship: the speaker influences the attitudes, beliefs and/or actions of the listener. Is persuading merely a special subcategory of informing? If so, then the first stage is tautological, with the goal being to persuade, and the other three stages are subservient to this dominant purpose. Messages are created, presented, and evaluated as a *means* to accomplishing the goal, rather than providing various aspects of information transmission that must balance a number of, possibly conflicting, purposes. Here, the emphasis remains on the strategic selection and implementation of those aspects of the communicator's craft most likely to move the listener(s) where the speaker wants them to go. Contexts where this is most apparent include campaigns to sell people or products, negotiations, sales transactions, complaints and criticism, and decision making in the myriad of situations and institutions where people must collaborate.

How to persuade? Since one has to "move" the listener rather than simply augment her or his store of information, strategy must be grounded in analysis: discovering who they are, "where they're coming from," their point of view. Listeners possess a specific set of values and beliefs, they have certain needs, and they may already possess (or be encouraged and inspired to develop) certain goals or expectations. Speakers try to apprehend all of these aspects of the listeners' world, and to develop and employ a strategy, at least consistent with, if not based upon, them. While reason and rationality may get you through the door, psychological and motivational appeals are much more likely to gain you an invitation to dinner.

Perhaps the effective speaker's foremost quality (here as well as in the framework of information-giving) is credibility. To sell something people perceive you yourself don't believe in is a herculean task. A second vital quality is the ability to develop rapport: when you have listeners empathizing and identifying with you, their inclination to contradict and struggle with you is undermined. Many strategies to establish this human connection exist, and among the more important fall in the nonverbal realm: tone, voice, pace, gesture, eye contact. As speakers attempt to persuade and convince, they use language as a tool and must master it; but knowing the material they are working with is as important, and when that material is the consciousness of other people, it is constantly in motion.

As to content, the context will dictate. In the light of pure reason, we need only develop the perfect argument, marshal supporting evidence, and carry away the prize. In the diverse real contexts where persuasion dominates, many additional choices arise and much more judgment is needed. However subtly, the successful persuader is usually in control of the situation. But often, to be successful, such control must be accomplished with velvet gloves. Knowing when to pull and when to slacken the reins involves both the choice of words and phrases, and a keen reading of the listener's developing attitude to the subject at issue. In the end, though many excel at persuasion naturally or intuitively, the game is psychological.

Conversely, the listener can best fortify her or himself against the process of persuasion with their awareness of these psychological appeals. Separating the message from the speaker will immediately blunt the force of much persuasion, allowing the listener a "piece" of mind for each of two conflicting desires; to like but yet to disagree. This strategy ensures that the listener will be exposed to as much information as possible or desireable, while reducing the risk of being unduly swayed. Becoming conscious is the key, conscious of: the line of argument, the supporting evidence, the relevance of one to the other, the facts (as opposed to everything else) and—obviously—of all of the subtle appeals being employed. Credibility is the listener's to grant or withhold. Values and self-awareness have much to do with resisting the appeal of others. To fully open one's mind and heart up to—and then to resist—the appeal of another is a significant transaction, not a failure to communicate.

Relationships

Information giving and receiving as a river, persuasion as psychoanalysis, speaking and listening as a dance—all of these metaphors point to the dual nature of the subject. Group 4 author Stephen Witte wrote of reader and writer intersecting at the text sign, where meaning resides. With speaking and listening, Group 3 believed, an even greater emphasis must be placed on the relationship that exists between the communicator and audience. More than mere context, or another important parameter of communication, in this realm relationship *is* communication, and the better the one—whichever—the better the other.

The group's view of relationship grows out of—and mirrors—much of the earlier discussion. From this, it is clear that the listener's role is anything but passive. Nonetheless, it is usually the speaker who is first to *experience the consequences* of a relationship failure, because her or his goal was thwarted thereby. Listeners of course suffer from lack of knowledge, from not having been "moved," but their awareness of the failure can be less acute, depending on their already-present motivation to receive the transmission. Assuming they *are* interested in the speaker's offering, they nourish the relationship by encouraging it and supporting it in any way possible. Speakers (being human) like to be liked and approved of, and thus the more involvement the listener manifests, even if it is only demonstrably empathic listening, the better.

A few elements illustrate this dynamic. In a speech act, listener and speaker are interdependent; clearly, neither can function without the other. Often this mutuality can become manifest when the situation is framed as a group problem-solving exercise (even if the problem is nothing more than successfully conveying information). Listeners can be co-creative by assuming an overt role of responsibility for the success of the communication encounter. The goal is a shared vision and attaining it becomes a mutual or team enterprise. The speaker is providing the fuel, and the listener(s) must encourage the event in any way possible. Using verbals and nonverbals, listeners generally minimize their ego in the encounter, expressing themselves actively only in order to assure the encounter function better.

Certainly it is the speaker, however, who most often keys and styles the communication relationship. Group 3 identified a handful of principles, and strategies to implement them, of which good speakers need be aware.

○ Communication is a process, and good speakers are continuously managing both the conversation and the larger process of the relationship. They take responsibility for staging the encounter in a friendly, unintimidating way. Where assertiveness is useful—and it can often be functional and liberating to have a leader in any group process—they provide it. But involving the listener(s) is crucial, and the speaker is responsible for getting this interaction in motion. Speakers must attain and maintain a meta-view of the communication encounter. Decisions must be made continuously—based on vigilant surveillance and fine-tuned awareness—about any problems that arise. Those that do fall to the speaker to at least devise a solution to, using a number of strategies. The group noted that a substantial body of research on problem solving exists and is relevant to this arena.

○ As mentioned above, speakers can provide a sense of "groupness" and thus draw the listeners actively and productively into the communication frame. A basic strategy is to establish rapport. The seed of the idea is planted that *we* are doing something together, and speakers in subtle ways convey to the listener(s) this mutuality. A direct way to do this is to identify common interests, concerns, and background. More indirectly, speakers use rewards, compliments and nonverbal cues to encourage the listener's participation. A good speaker will often be flexible about the roles played in a conversation, granting more apparent control to a listener when it seems that structure will overcome resistance to being controlled. Active participation by listeners unconsciously brings them into the frame, provides them a platform for developing commitment to the success of the group undertaking. Thus do speakers learn not only to present but to generate talk between the two sides of the relationship. Speakers can provide an excellent model of a listener by how attentive they are, how they respond to other viewpoints.

When the listener(s) are many, and the process really is more of an overt group being primed by a primary speaker, some of the above strategies become crucial, and more explicit. The goal is not only to generate commitment among listeners but to establish an interdependence that can eventually cement a consensus on the content of the communication. With a group of people, coalitions are constantly and naturally forming, breaking apart, and realigning. Here the group noted was another area where a great deal of research and analysis has been done. A number of participatory verbal group behaviors arise, as people establish positions, negotiate, assert themselves, ask questions and seek clarification. From these arise another level of behaviors one step closer to consensus: speakers and others learn to accommodate the interests of others, and heighten the stakes by sharing their own approach, disclosing personal information, and clarifying differences among the group in the aid of negotiating a compromise. A final stage is reached when the group can begin to talk about "us": arguments and positions are openly integrative, agreements are highlighted, the collective enterprise and a higher purpose is invoked, and the group's collective consciousness is finally synthesized and connected into one position.

How do speakers manage this group process successfully? A number of strategic moves can be specified, but all work better when seen as strategies to enhance the relationship, rather than to manipulate the listener(s). These moves highlight the interactive basis of the relationship underlying effective communication: effective speakers are attentive and responsive. They pay attention to everything and everyone, and also manifest clearly this attentiveness: people feel as if they are being heard and valued. This can involve a short list of common sense, respectful rules; politeness, empathy, understanding, appreciating and reaffirming the needs, interests, and points of view of the listeners.

When the "listening" group is more complex, numerous, and probably more fractious, more strategies may be needed. These are laid over the basics just listed, and here the speaker is probably more obviously orchestrating the dialogue. Listeners grant speakers this leadership status for various reasons, but being perceived as fair and credible are the most crucial. Again, the number of situations that can arise are many and complex, and again the group suggested that solid literature on this area exists and should be consulted. A very important skill is the director's (speaker's) awareness of the timing of human talk. As discussed earlier, the cognitive and emotional substrate which precedes and "times" the back and forth of conversation is complex to analyze, but fortunately most have a good sense of it, and a group speaker/leader must be exquisitely aware of this element of the conversation, not only timing his own delivery and presentation but helping others to "be heard" and controlling others who would interfere with the kind of smooth flow that invites everybody into the stream of discourse. Figuring out how people need to take turns, when they should be interrupted or brought gently back to the point, how much digression is permissible or necessary, and how much integrating and summarizing must be imposed onto the discussion are some of the more central issues of concern.

Finally, the speaker needs to arrive successfully at her destination. The speaker initiated the entire encounter, provided most of the content and direction of the information, and had a goal or purpose. While the end is too late to recoup crucial lost opportunities, it is exactly the time to ring home the final note whose echo will remain with the listener. So much information, so little time, is the quintessential dilemma. And in most communication acts, much more was obviously offered, delivered, and received than the main point. Nonetheless, there usually *is* a main point, a central vision, and the finale may be the time to recreate that vision in its most powerful and simple form. Whether this involves summarizing, referencing key points, recalling highlights, or moving major sections of the conversation into the framework where they will best accent the goal, the speaker must conclude, and the relationship on which the communication was based is likely to have determined how successfully the goal was met.

Conclusion

Group 3 thus was able to produce a fairly rich catalog of the abilities involved in speaking and listening. Efforts to shoehorn these insights into the NCES icon were sincere, but ultimately frustrated by the group's belief about the layered complexity of this type of communication. The model they did develop, as outlined above, they hoped could provide planners and even testers with a robust framework to proceed. But while they were able to follow the NCES charge more closely than their colleagues in the reading and writing group (Group 4), they were no less convinced that assessing this type of higher order communication ability was a precarious undertaking, absent a very sensitive awareness among those assessing of the many confounding factors. If relationship is so crucial to the communication act, can it be fairly evaluated in snapshots? Probably not. However, given the generally low—many had no hesitation in describing it as woefully inadequate—state of higher order speaking and listening not only in the colleges but in the populace at large, the group felt their model provided a productive jumping off point. They did agree with Group

3 that this was by no means an overnight process, and thus hoped that their insights about communication could provide some important foundation.

READING AND WRITING, GROUP FOUR - AUTHOR - WITTE

PARTICIPANTS

Facilitator:

Steve Hunt
OERI, Office of Research

Author:

Steve Witte
University of Wisconsin, Wisconsin Center for Education Research

Reviewers:

Joanne Carter-Wells
California State University, Fullerton, Dept. of Reading

Michael Galgano
James Madison University, Dept. of History

Marcia Farr
University of Chicago

Participants:

Richard Battistoni
Rutgers University, Civic Education and Community Service Program

Robert Chalmers
Purdue University, Dept. of Pharmacy and Pharmacal Sciences

David Elias
Eastern Kentucky University, Dept. of English

Ellen Porter Honnet
Johnson Foundation

Dolores Lipscomb
Chicago State University, Academic Development

Marty Nystrand
University of Wisconsin, Madison, Dept. of English

Sharon Rubin
Salisbury State College, Dean of Liberal Arts

Pete Smagorinsky
University of Oklahoma, College of Education

Others:

Joan Eisenberg
Recorder, George Mason University

Cliff Adelman
Office of Research

Otis Elliott
Fund for the Improvement of Postsecondary Education

David Loope
National Center for Education Statistics

PROCEEDINGS:

Introduction

Group 4 convened to discuss assessment of reading and writing in the light of author Stephen Witte's paper and the attendant reviews. Facilitator Stephen Hunt thought the deliberations remarkable in that people from a wide variety of academic disciplines and backgrounds almost unanimously endorsed Witte's underlying theory of communication, as well as his apprehension about an instrument-based assessment. And so it is upon the foundation of some of Witte's defining concepts (and not the NCES icon) that the following is based. Notwithstanding Hunt's characterization of the Group 4 deliberations as lacking diversity and dissent, this choice of organization should not be mistaken for a mere gloss on Witte's paper. He will himself provide that viewpoint, in a revision of his paper based on the Group's discussions. Rather, the framework is useful because it is holistic and contextual, and (as Hunt emphasizes) the group shared Witte's theoretical perspective. Thus they were able to proceed to a wide-ranging consideration of how-to design, and even implement, an assessment, a task which could have been, because of their shared emphasis on context, extremely esoteric to grapple with. By essentially adopting Witte's eight major "skill sets,"[12] they were able to proceed to what they felt were some of the more important details, details they believed should drive the form of the assessment.

For a couple of important reasons—as Hunt emphasizes—the NCES icon failed Group 4 as an organizing device for their discussion. First and foremost was the fear of reification, of losing the overarching significance of the contexts of larger production and use when "skills" are itemized and isolated for assessment. In fact the group preferred the term "abilities" because it better conveyed the constructive and dialogic nature of good communication. Second, was the related feeling that simplistic identifications of separable [and only arguably so] aspects of reading and writing might lead to a group product that, in Hunt's words, "could be used to create the variable elements of a test instrument," an instrument of the traditional sort which stands in opposition to the centrality of dialogic and constructive events that lie at the center of all communication activity.

The Nature of Reading and Writing

What are the higher order abilities of reading and writing? This framing question led Group 4 to produce not a list but more of a credo, at the heart of which were not discrete, analyzable skills, but rather a sets of abilities that enable communicants to engage in a process. A process which involves:

- The ability to perceive and appreciate the context of any communication;
- The ability to recognize and act on the awareness that all communication is a form of dialogue, and the ability to identify with whom (and how) to continuously collaborate in a socially constructive dialogue;
- The ability to interpret and use a choice of language, symbols, meanings and intimations that are appropriate to—and constructive in—the contexts of language production and use;
- And finally, with all of these controlling frames identified, the ability to orchestrate one's abilities and the process itself skills to accomplish the goal(s) of the communication, in the context of a social act.

This credo identifies the hallmarks underlying effective communication, what might be established as the target of reading and writing assessment. But the practical realities of American civic and workplace life, as well as the current state of postsecondary students' abilities—in the light of this target—present enormous problems, because effective communication is *not* everywhere to be found. Group 4 saw their task as answering this question (How do we try to reconcile these problems?), and this task seemed to contradict the charge given them by NCES to develop a list of skills. Their answer as to how to reconcile the problems was, in fact, immediate and forceful, and their deliberations were spent mainly on working out the implications of it:

> ***Do not devise a quantifiable, ETS- or ACT-type instrument for assessment.*** If you do, the baby will be irretrievably lost with the bathwater, because the holistic, social, performance-based, context-dependent essence of effective higher order communication can be elicited only: in authentic situations, over a period of time, and based upon extended writing samples that demonstrate continuous and active construction of the social context in which the communication will engender a dialogue.

Manifestly, and unfortunately, this definition of higher order communication does *not* describe the skills possessed by most college students in America, nor does it fit very easily into the infrastructure and implicit values to be found actually operating in most universities. While this gap has been continually described and bemoaned by many in the assessment dialogue, often it has been followed quickly by a pragmatic corollary, which runs something like: "Well, it's too bad our students and universities aren't *there*, but our job is to see where they actually *are*, and to devise an instrument to describe them *in situ.*" Group 4 rejected this sort of response, because of the earlier-described paradox that stems from the nature of communication.

Higher order reading and writing, they believe, simply cannot be measured *at all* by a machine-scorable, objective instrument. Any instrument designed to target this complex of abilities will fail. Ergo, don't develop an instrument. Develop, instead, the beginnings of a new operating system within the university, founded on the process nature of higher order reading and writing, a system characterized by the irreducible elements of context and complexity, one reliant on evaluation of extensive writing and revisions, over a sustained period of time, and framed by the social ways in which people collaborate in the act of communicating. They foresee that such a system could evolve, even as the skills of the students improve, and as faculty come to invest in and embrace the institution-wide nature of the challenge.

Group 4 was not engaged in a visionary or utopian exercise. Rather, as their discussion clarified the process nature of higher order reading and writing, they were able to point toward how an individual school could begin to reorganize around what should be no less than a primary ingredient in its purpose. Such a system would begin a far-reaching and revolutionary restructuring of university life, however, and the group was not naive about the task. An important first step, they believed, was to clarify how and incidentally why the process of reading and writing would be a fundamental prerequisite to such a restructuring. A number of themes and interconnections emerged during the discussion, and a number of keywords point toward them: faculty participation, context, cultural diversity, multiple levels and assessors, and capstone event. [A chart or mosaic would probably better capture the discussion than a sequence of points, but the latter lends itself better to the future dialogue.]

Context and Diversity

What does the Group 4 mantra of context actually mean? The relevance of context as a defining framework applies to several levels of the postsecondary picture. To reinforce their dissatisfaction with an instrument, the group emphasized that testing situations provide a context *ipso facto*. "Reading" this testing situation and succeeding in it are both valuable abilities. But they are not necessarily relevant to the role reading and writing have in the workplace nor even in courses (of which the group approved) that involve extensive writing, critical research and revision. The testing context is—if not false, then—misleading, and it has been found to perpetuate cultural biases and inequities.

A much preferable approach is to identify and try to establish authentic situations where: (a) context actually defines and calls for a particular communication strategy, and (b) permit these situations to be actively modified by the communicator. Higher order abilities are manifest when the communicator must bring judgment, analysis, and orchestration of this situation into play. Authentic situations in the world after college are unlimited, and thus a crucial communication ability is learning how to learn about the exigencies of this unending stream of new contexts. The multicultural nature of many schools, and the increasing diversity of American civic and workplace life put this ability at the heart of good communications. A further implication to be drawn from the central role of context in communication and the multicultural nature of American universities is the need for multiple solutions: multiple assessors, multiple levels of approach, multiple frames of assessment.

Narrow notions about the social, holistic elements of communication betray a metacognitive flaw in the communicator, an inability to realize that communication occurs in context, and that each context comes furnished with useful/appropriate functions, uses, roles and purposes. One way to describe this skill is "mapping the situation;" or developing one's internal representation of all the cultures relevant to the situational moment.

Faculty and the Curriculum

How does this view of communication intersect with faculty at the postsecondary level, with what and how they teach, and with the reading and writing they expect from their students? Except in general education or developmental reading courses, the process and constructivist nature of reading/writing is the exception rather than the rule on American campuses. This is so for many obvious reasons, the lack of resources and thus a larger class size being perhaps the foremost. However, a student pursuing a major does come to learn the rules, the jargon, the procedures and the philosophy of a distinct discipline. This context may be a student's most intense *application* of the lesson that context is crucial, for each field has its own elaborate and undeniable context, mastery of which is necessary to do good work. And progress, defined as advancing the knowledge of a discipline, will require students who go further to reconstruct the context in which they are being taught.

But there is another way of framing the problem. Are faculty within these distinct departments unable or unwilling to convey their lessons as a particular case of a general skill? That is, to frame, generalize and articulate a metadisciplinary view of what they are about, which would provide their students with a metacognitive introduction to the concept of good communication? Obviously some are and some are not. Some do and some don't. As long as such metaphysics is left to the inclination and skill of the professor, student bodies will receive and perceive only sporadic and haphazard links between what they are doing in their major and the more general activity of learning to read and write in the service of effective communication. When portfolios are developed and evaluated during the pursuit of a major, so much the better. But the mere collection of portfolios doesn't draw out and develop skills that underlie good communication, skills that could be assessed only over a period of time, with logs, journals, observations of process, and ultimately self-reflection, evaluation, revision and correction. Very few schools (Alverno College is the exception) have established the comprehensive, interdisciplinary infrastructure where such practices can be meaningfully encouraged and assessed. But for the often cursory involvement of an advisor or counselor, most students have a series of insulated encounters with one after another teacher. Continuity of their learning experience is up to them, but surveying and managing this continuity does not come automatically. At Alverno and a few other institutions, faculty bring a multiple view to the individual student, providing a far more holistic system of effective feedback which provides a continuous reconstruction of their dialogue.

The Alverno model points to another element the group saw as crucial, the interface between postsecondary life and the workplace. If communication is so reliant on a sensitivity to context, the danger exists of developing "successful" students who master the concept in college, but cannot transfer that mastery to subsequent workplace settings. This provides another warning about assessment instruments and the dangers of overlooking the cultural diversity of American students. Mirroring the changes in society, schools are no longer (if ever they were) factories that churn out a homogenous product. Much greater diversity of race, culture, gender and age now defines the postsecondary population. These people are different before they arrive on campus, their experiences there develop differently, and they emerge looking at different ways to employ their education.

The differences are too great to homogenize. They must be incorporated into a more inclusive view of education. This new frame will not be erected on a list of skills that subserve the delivery of a so-called universal message. There is no universal, underlying core to communication: it is social, constructive and interactive, and can only be defined (its abilities described) in the act itself. Most importantly, to measure communication is to measure effectiveness, a property which tends to emerge with numerous perspectives attached, not susceptible to a numerical score. The whole is greater than the sum of the parts.

Many of these perspectives can be labelled, however, and can be adapted to the content of most of the postsecondary curriculum. Especially so if the courses themselves begin to emphasize two elements that—admittedly and problematically—are more easily incorporated in smaller, seminar-type situations. First, that learning happens as a dialogue, between teacher and students, and among groups of students; knowledge and meaning are not reified products, but rather a continuous mutual construction. The second element follows from the first: values and beliefs underlie all communication; only through social discourse, argument, revision and evolution does consensus emerge. This view does not slight the importance of the irreducible intellectual *content* of coursework and disciplinary bodies of thought. Rather, following Witte, it emphasizes that any such content is subservient to its uses. Such a content cannot be *transmitted* to people, from on high as it were. Rather, a systematic dialogue between the information (perhaps as embodied in the teacher's presentation) and the learner ensues, where learners must articulate their own comprehension of reading material, through analytical conversation, writing, critiquing, and revision. The group thus began to specify "authentic learning experiences," which could be promoted throughout the curriculum.

Understanding—learning—ultimately comes to reside in the student through the process of construction, but it cannot be injected like some growth hormone. Understanding is a constructive activity, an exercise, and only by training and developing a student's muscles of awareness, discernment and expression can it be perfected. Thus, faculty would learn to engage students in the exercise of learning through the use of dialogue: at times Socratic, at times directive, at times argumentative and probing, where possible developing a menu of case studies, laced with creative doses of analogy, metaphor and a balancing of new and given information. If such a model of instruction were somehow encouraged, rewarded, demanded, group members were generally optimistic that faculty would come to embrace it. Such a rethinking of the curriculum could lead to a more far-reaching revision of postsecondary education. The group did not think it naive to operate on a hopeful premise: the system will move, if driven by the palpable improvement of teaching and learning. Implicit in this evolution is a broader and deeper view of the teacher's role and ability to inspire and provoke the spark of learning, their specialization notwithstanding. This broader role favors the multiple assessor approach, as groups and teams of faculty are more likely than a single individual to produce a deeper awareness and a richer evaluation of a student's abilities. The question then, is how to engineer this revolutionary turn in the road.

Assessment as a Capstone Event

First and yet again, to reject categorically the option of instrument testing. If assessment of reading and writing is to be done, it must be done "*correctly*," even if that means imperfectly, piecemeal, and on a less than national scale. The group's premise was that good ideas will grow. Thus, an important first step is to collect as many as possible of the innovative ideas that have already been developed. (A partial list: Alverno College, Evergreen State, Vermont and other portfolio experiments, Assessment forums like AAHE and AAGS, James Madison U seminar paper/public reading day, etc.)

Next, devise an approach to learning and instruction that will serve as a symbol for the entire learning experience. If the process and meaning-constructive nature of reading and writing is fostered and expected at all levels of college, it begins to become a metacognitive style of learning. The concept need not be vague or amorphous, however. Developmental reading is taught as a higher level skill. Critical thinking courses have begun to flourish. In this model, all such components would be subsumed as general preparation for a capstone event, which would probably be spread over at least a full semester of the senior year. This course would then culminate the experiences developed throughout college, and, thus, provide a basis for assessment with maximal validity.

A good place to start would be to establish a pilot program, intended to develop a model that could be offered to a wide range of schools. As the capstone course revealed both the strengths and weaknesses of students participating in it, this information would then loop back in a schoolwide process of evaluating teaching methods. While the two would inform each other, the group insisted on the integrity of the core elements of such an event or course. It must proceed through extensive written products. These documents must be considered as a protean sort of template, constantly under revision. The documents and their authors must be tested by challenging, external audiences, various audiences, one capable of expert criticism as well as another, ignorant of the subject, in need of persuasion. The course must balance individual product and group collaboration. It should express the advanced content and sophistication of disciplinary work, yet require that such information be transferred and delivered into other contexts.

This transfer or translation element, the group believed, could provide a dramatic opportunity to factor in the workplace and citizenship issues posed by Goal 5. Both the content and the method of such a course could be inspired by rather more "practical" than academic situations, anticipating the demands students will likely encounter at work and in the community. The relevance of such case studies could be continually checked through feedback and guidance from the world beyond the university. Many possible avenues of dialogue for constructing and revising the content of the course could be developed (professional associations, corporate trainers and recruiters, professional personnel people, local community activists, etc.).

In this way, the capstone event becomes a fulcrum. It epitomizes, accentuates, and evaluates the concept of learning to learn, and does so in an authentic situation, constructed around social and group processes. It allows for the kind of extended revising and collaborating that more accurately reflect the constructive process nature of reading and writing. And at the same time it can serve as a true commencement, a transition to the multiplicity of contexts students will begin to encounter after graduation.

Is it practical? Is it feasible? These questions would seem to revolve around three crucial elements. First, the cooperation of the schools. Second, the resources and directive of the political voice. Finally, the will to persist, holding to the highest possible conceptualization of higher order reading and writing. All three of these issues are problematic, to put it mildly. Faculty "buy-in" to such a far-reaching shift of responsibilities and roles, in many situations, would involve no less than a major revolution. The practical, political question of strict accountability, the search for a "quick-fix," the motivations of the various actors all would need to be overcome. But it is the underlying value of the perceived results of the experiment that will have to drive these other two. If American students can begin to rise to a level where higher order reading and writing is an essential vehicle for their moving through college, then a capstone event could provide a dramatic showcase to faculty and the world beyond of the university's vital role in society.

JO ANN CARTER-WELLS, Department of Reading, California State University, Fullerton (Group 4, reviewer)

[These remarks were excerpted from her review of the Witte paper]

Concerns, Issues, Clarification

There are a few concerns/issues/clarifications that have grown out of this paper and previous readings. Primarily, there does not seem to be a clear construct for "communication skills" in Goal 5.5 nor in this paper. A particular definition of postsecondary reading and writing would be helpful. "Advanced ability" is an obviously problematic term as clearly identified by the author with a lack of supporting data in *America 2000* (1991). There are many assumptions about the necessary communication processes for success at the postsecondary level. Chief among these is the list provided in the "Green Book" on Academic Preparation for College (College Board, 1983). In reality, these lists reflect neither the sophistication that actually exists in different disciplines, the complexity of the processes, nor the developmental differences (Carter-Wells, 1988). In addition, the outcomes literature has not provided an in-depth understanding of the growth in communication skills. They are assumed to be developed through a composite of academic experiences. Indeed, even critical thinking has been studied using "narrow" instruments such as the Watson-Glaser Critical Thinking Appraisal (Astin, 1978; Bowen, 1978; Pascarella and Terenzini, 1991). Finally, learning theory as it relates to adult developmental theory are relevant components of all these processes that do not seem to have been articulated in any concerted way to date in this movement.

Another issue is the emerging linkage of the critical thinking assessment movement to both the workplace and citizenship. Postsecondary reading provides a strong background to these areas. Many workplace literacy leaders/instructors have come from the ranks of college reading/learning. The emphasis on "learning to learn" skills/processes are the underpinnings of any credible maximalist reading/learning curriculum where students become decision-makers taking responsibility for their own learning (Mentkowski, 1991). The recent election demonstrates the necessity of including the reading process in citizenship decisionmaking (Morse, 1989). Consider the propaganda in one's mailbox or the ballot pamphlets with the initiative analyses, arguments, and rebuttals written at a level of sophistication far beyond the average voter!

Finally, the value or affective issues in this movement are paramount in not only a multidimensional conception of critical thinking/communication skills assessment but also in teaching and learning implications.

Assessment

The calls for assessment of the critical thinking/communication skills in a multidimensional framework has the potential of profoundly influencing the assessment of the individual curriculum domains. In the design of any future instrument(s) related to this national effort, consideration of the readability of passages, vocabulary complexity, concept density, disciplinary differences, and academic/developmental levels of the test takers is crucial. Extensive research has indicated high correlations between traditional college reading and critical thinking measurements. And verbal reporting data indicate vocabulary confusion and creative thinking beyond the domain of these instruments (Carter-Wells, 1992; Hopper and Carter-Wells, 1988). Are we testing reading comprehension or critical thinking or both at the same time?

Research opportunities abound with this assessment design not only with the correlation of critical thinking and communication skills in terms of construct and content validity but with underpinning issues such as gender differences, "Ways of Knowing," adult development, reflective judgment, disciplinary differences, the influence of required curriculum (such as in California), existing placement testing, the role of creativity, etc. (Lynch, 1992; Magolda, 1992).

Teaching and Learning Implications

Separate from any political and/or budget issues are two looming teaching and learning implications. One is a teacher training issue in that most instructors in this area have not had formal training and/or certification. There are very few Master's and/or Ph.D. programs specifically devoted to any of these areas. Self-training appears to be the norm. This has implications for continuity and currency of related instruction in relation to reported assessment data. We have seen this same phenomenon with the data from the NAEP at Advanced Reading levels. Recent interest and research in verbal reporting with peer discussion and peer editing activities is only one example of the type of application activity that needs to be incorporated in teacher training programs.

Another issue is the importance of critical thinking and communication skills in the postsecondary curriculum. Currently, these are usually packaged in General Education. If these are to become increasingly important outcome areas, then disciplinary/major inclusion will be necessary. Innovative accreditation agencies such as the National League of Nursing are influencing total curriculum implementation. The parallels between the literacy and critical thinking movements are striking. *Becoming a Nation of Readers* (1984) may soon be expanded to *Becoming a Nation of Thinkers*!

Critical Reading/Critical Thinking Skills

There could be some disagreement of the simplicity of defining "basic skills" as incorporating the processes of production and comprehension, especially among the literacy community or those developing statewide entry assessment testing. However, the conceptualization of "advanced ability" includes important underpinnings that link critical thinking and the processes of reading and writing in the postsecondary/career experience, namely, metacognition, decisionmaking, advocacy of one's position, linguistic flexibility, prior knowledge or schema, and disciplinary differences.

However, for purposes both of instruction and assessment, we have developed a list of skills that reflect these processes. This list is an integration of critical thinking and college/adult reading based on the Delphi consensus agreement definition of critical thinking (Facione, 1990). They also represent the most recent theoretical perspective in the field of college reading and learning (Lewis, 1991). It is interesting to note that many of these skills are also included in other lists presented both for this workshop and that of last fall from two other disciplinary orientations (Halpern, 1992; Paul, 1991).

INTERPRETATION

Constructing tentative relationships among ideas, e.g., main ideas and subordinate ideas.

Constructing a tentative organizational pattern for the text.

Distinguishing important ideas from unimportant ideas.

Clarifying meaning of words and phrases.

Paraphrasing ideas in one's own words.

Hypothesizing meaning and reading to confirm or disconfirm.

Inferring meaning by integrating information from the text, from prior knowledge, and from evolving interpretation.

ANALYSIS

Reflecting upon and embellishing interpretation.

Generating author's purpose, theme, point of view when not stated.

Identifying mood and tone.

Constructing issue, main conclusions, and reasons.

Generating unstated assumptions.

Distinguishing fact from opinion.

Tracing the relationship of paragraphs to each other and to the aim of the discourse, point of view, main conclusion, or theme.

Recognizing bias in language.

Conjecturing implications of ideas, e.g., results of accepting or application for use.

Reconsidering tentative organizational pattern and constructing names for the parts, e.g., situation controversy grows out of, author's position, reasons, alternative solution.

EVALUATION

Assessing the credibility of the author and the source as well as authorities cited in support of a claim.

Judging reasoning used in interpretation and analysis.

Assessing alternate interpretations.

Judging acceptability of author's ideas.

INFERENCE

Querying evidence to construct meaning, conjecturing alternative meanings, and drawing conclusions necessary to interpret, analyze, and evaluate.

Deciding on best plan for reading a text for a specific purpose.

Constructing a solution to a comprehension problem.

Seeing significant implications of the advocated position.

Identifying and securing information needed to form a synthesis of multiple sources.

EXPLANATION

Summarizing or constructing a graphic display (graphic organizer, semantic or concept map) of the results of the interpretation and analysis.

Explaining criteria used to judge quality of interpretation.

Explaining how an interpretation meets the criteria of consistency, relevance, completeness, and simplicity.

Advocating a specific interpretation.

Explaining criteria used to judge acceptability of author's ideas and how ideas meet the criteria.

Comparing ideas between reading sources and explaining similarities and differences.

SELF-REGULATION

Monitoring comprehension.

Separating one's personal opinions and assumptions from a writer's.

Varying reading speed and method according to the type of material and one's purpose for reading.

Choosing and implementing strategies for circumventing a comprehension break down.

Considering possible alternate interpretations.

Knowing when to implement different levels of thinking, e.g., when to move between interpretation, analysis, and evaluation.

It is difficult to give examples that fit only in one category. This is not the result of a flaw in the definition of critical thinking, only a reflection of the complexity of critical thinking and reading. Comprehension construction is a recursive process involving inference, interpretation, analysis, evaluation, explanation, and self-regulation.. The first words encountered are external stimuli which send the brain on a cognitive search for relevant information from which meaning can be inferred. The first piece of meaning is an hypothesis to be confirmed or disconfirmed. As the process continues, adjustments are made through analysis, explanation (to one's self), evaluation, and self-regulation.

Also, new interpretations are added to the evolving version of the text through integrating relevant information from the page, prior knowledge, and interpretation previously constructed. The point at which a specific kind of thinking is employed is unique to the text and the reader. For example, analysis may be employed long before the passage is completed. This is based on the theoretical framework of the prediction capabilities of fluent readers.

The results are similar to a first draft in writing. The first draft is revised or extended through additional processing of the text or reflection upon the reader's mental representation of the text. All drafts are interpretations, but all interpretations are not the same. The quality of the first draft depends upon the reader's need, effort, and ability in employing cognitive skills of critical thinking. The ideal outcome is the most sensible, coherent, and justified version.

Finally, the concept of creative thinking is also employed in critical reading curriculum when readers extend beyond the author's ideas in a variety of ways. These include, but are not limited to, the production of a unique communication, plan, or abstract relationship based on the extension of their interpretation and would reflect differences in prior knowledge, developmental and academic level, and disciplinary differences.

Part 3. CLOSING COMMENTS

FINAL OPEN SESSION, INDIVIDUAL COMMENTS

ERNEST SAVOIE, Director, Employee Development, Ford Motor Company. (Group 2)

Thank you very much. My name is Ernie Savoie; I'm with the Ford Motor Company. I'm Director of the Company's Employee Development Office. We have about a dozen employee education and training centers at Ford, so in a sense I'm also the "dean of education." Because of these education and training centers and because of some of my other employee development functions, I'm heavily involved in the transformation of Ford in terms of quality, continuous improvement, customer satisfaction, and labor relations. In recent years, we've had an excellent record across all these areas.

I thought our group was particularly effective. I was with the critical thinking, problem-solving group, although I don't believe it's possible to separate out from these subjects in the matter of communications. All three of these subjects come together in real life.

I thought we had a good give and take, much better than in many of the Washington groups I've been associated with. There was true listening by all members of the group; there was real sharing of views and experiences; and we made considerable progress in not only refining a model but also in applying it to the three things we talked about: the worlds of college, work, and citizenship.

I believe that the Critical Thinking Project may have important implications for employers in three ways.

First, the findings and recommendations can help the schools deliver a better product. If the schools and the students learn from what we discover, and if they use that knowledge to create a better education product, employers will get better employees–more competent employees. That will help improve American competitiveness.

Second, the effort may help some employers–particularly large ones–if they can adapt the findings and use them as tools for individual and organizational development. I can already see how we probably need a broader course in critical thinking, for example, before we go into applied problem solving.

And third, under the right conditions and circumstances, employers may be able to directly apply the findings and recommendations to people assessment. I don't mean as a "go/no-go" proposition–a means of deciding whether or not we hire job applicants. For one thing there would be legal ramifications, because what you're testing has to be work-validated, and validated principally for entry-level jobs. We can't argue: "Well they're really going to need the skills we're measuring when they get to be managers, some day." We'll have to be certain that our ideas are submitted to those people who have expertise in the legal ramifications of selection and job-relatedness.

Another concern about linking measures of critical thinking to assessment is that so many other factors impact success. They go far beyond just the rational and measurable. You get into courage, risk-taking, economic conditions, and other outside factors. All of these will help determine the success an individual achieves; it's not just the critical thinking factor.

Whatever we try to fashion regarding how we evaluate critical thinking has to be carefully thought out with respect to actual problem solving. We talked about that in our group, but we didn't reach any hard conclusions. We have to communicate to those who will be developing measurements that a lot of problems take a great deal of time to analyze and solve. You can't just cite four, five, or six possible applications of problem solving and measure the immediate responses. Understanding and factoring in the time dimension of problem solving is very important. So is an appreciation that the really significant applications will vary from employer to employer, and from situation to situation.

The kind of problem solving/critical thinking that our group talked about was for age 20 to 24 college-level people. So, I think it needs to be made very clear to employers that that is the order you're assessing, and not higher levels of critical thinking that employees might need later to perform at managerial levels. The latter levels would require different forms of assessment and would probably require factoring in post-school experience.

In addition, we would want to stress with schools, with employers who would be using our measures, and with the students themselves, that there's a "learnability" aspect to critical thinking and problem solving. You have to learn higher order thinking all through your life. You're not done because you've been assessed and somebody says "Hey, you're pretty good at higher order thinking." The idea of continuous improvement in learning ability has to be embedded in our thinking, and the results of an assessment at an early point in your career doesn't mean you've got it or you don't have it. It has to be a process of building on.

This means we'll have to guard against the development of overly simple instruments. And I would hope there would be richness and depth. I know we can't do deep assessments for a million students because of time and the cost involved. I know it has to be done on a sample basis. But perhaps, in addition to our simpler measures, we can have experiments where there is some deep stuff going on, and then we can try to aggregate these richer dimensions. I believe that would make sense to people. We would not just be relying on one approach.

We had considerable discussion in our group about the general abilities involved in higher order thinking. But there is always the question of how critical thinking actually works in the disciplines and the so-called applications–in marketing, engineering, labor relations, and finance, for example. The fear here is that if we come out with just a general instrument, a lot of people will not be able to apply it to their particular organization or their particular specialty. My view is that we should have sub-instruments that can be applied separately to the various domains or specialties. It may be that Part I is an instrument for general assessment, and Part II is for people in the various specific domains: biology, sciences, engineering, general business, specialized business, and the like. That way you could give realistic examples for those pieces. I despair of giving realistic examples that everybody, at age 20 to 24, would be able to tap into. I don't think the experience base is there.

I believe, incidentally that employers are as interested in the applications to citizenship and schooling as they are in applications of critical thinking to the world of work. Whoever designs the measuring instruments should test employers in these areas also. Half of what we can do today depends on what is going on in society, so we are very much interested in how our people react in citizenship areas. Let's remember that business is a customer of the citizenry, also. I think we need to reinforce that.

One of the results of this exercise could be stimulating a greater collaboration between the first 5 years of employment and the world of schooling. During those first five years, it's extremely important that focused learning continues. Critical thinking and problem-solving could be enriched by blending them into the realities of work. Employees and employers alike could learn, "This didn't work, but that did." That would be a real application in promoting critical thinking and problem solving.

Two other quick thoughts. While the focus of our efforts is on college, the employer clearly is also interested in the same types of skills at the blue-collar level, the community college level, and at the Ph.D. level. We have people functioning at all of these levels, and so we're interested in having critical thinking and problem-solving at all levels.

Second, if the assessment instrument is to be a tool–which I would like it to be–for individual development, it will have to go beyond the sample basis to establish validity for guiding individuals. So, I'd have to feel confident, when I pick up the tool, that when applied to Jim, or to Jane, I could really use it as something to increase their individual proficiencies. It's not enough to simply say, "Well, in general, everybody's about at an approximate level." I'd be hoping for an ultimate instrument that would serve us as a diagnostic tool for individuals and their personal development, and not just to be able to make general statements about the state of critical thinking in our college-age people.

Finally, many employers are exploring the matter of core competencies for the jobs their people perform. This involves making careful skills assessments of the tasks that need to be performed. If you looked at the core competencies for most jobs found in industry, you'd find almost everything that was discussed in this conference yesterday and today. So perhaps the people working on the critical thinking project should find out more about the core competency efforts that are already going on in industry. The two efforts may mesh nicely.

I didn't want to take that much time, but I wanted to share these ideas with you. And to thank you again for a project that I see as an important "can-do, will-do" attitude. If we focused on the obstacles, we'd probably just walk away from this and say, "It's too hard, and we'll never be able to do it." But what I detected is people saying "Despite the obstacles, we're going to do it." And that is a sure sign of progress. I thank all of you.

STEPHEN WITTE, University of Wisconsin, Wisconsin Center for Education Research. (Group 4, author)

My name's Steve Witte, and I'm an author. I have six points I'd like to address that may or may not be related. I'd like to preface it all by saying that I often see and hear more than I understand, and often say more than I should.

The first point that I would like to make refers to what came out of our group, which is a recommendation for a kind of capstone course, whereby a lot of different things occur. And one of the things that showed up very late in our session was that most people in the group agreed that it was workable. There seemed to be a commitment to pushing that forward.

The second thing that we found out very late in the day, today, was that there are a lot of examples around that should allow us to learn something about how to implement such a capstone course, where a large number of different activities are ongoing that involve—in our case—communication, but I think also not excluding problem solving and critical thinking activities. And that we need to have some way of pulling this information about implementation of similar kinds of things, some of which the government has funded. So that we're not reinventing things that we don't need to reinvent. And I guess that what I'm asking for is an appeal to the people at NCES to put together summaries of this kind of information so that in the future people have more to work with than we did.

Another point that I want to address is one of representation. And that has to do primarily with how what our group's—and I suspect the other groups'—work gets represented, after we all leave and go home. One of the things that we have learned is that each of the facilitators is going to produce a summary statement of what's happened in each of the four groups. I would like to suggest to Sal and Emerson that these summaries not be limited to in-house documents. Rather, that these summaries be shared with the people who were involved with the group. Because I think that the facilitators—as well as people like Sal and Emerson—need to be concerned with the question of whether the facilitator's representation of what happened is in fact shared with the people whose ideas are being represented. And that by the way is not to say anything against the facilitators. It's just a fact that things get represented in different ways.

And related to that, I think is the task that Addison has in front of him. I read with great interest, as I'm sure many of you did, the Summary Document [*A National Assessment of Student Learning: Getting Started*, NCES, 1992] of last year's meeting. And if you people were at all like me, you found that extremely helpful in terms of getting a handle on what had transpired before. I am concerned that Addison may not have the time to create the kind of thoughtful document he needs to create. And I would urge NCES to make sure that

Addison has the time to thoughtfully produce a document, that he not be rushed through that process. I try to put myself in his position and think, "My God, how would I possibly make sense of all of this?" I really would urge Sal and Emerson to support his cast as much as possible.

A third concern has to do with Penn State's involvement in this process. I spent time together with Sal and John Daly, and mostly I guess John, after Sal left last night, visiting with the Penn State people, and some of the apprehensions and anxieties that I possessed at the beginning of that conversation were somewhat relieved. But I think that I can't contemplate Penn State's involvement without continued anxiety. For much the same reason as just mentioned, that what Penn State's going to be doing is re-representing our representations of what we mean. And those of you who know anything about recorded sound know that digital sound is to music as squares are to circles. And that is the kind of loss of sound dynamics and sound quality that I'm concerned about. On the face of things, the Penn State group may very well say our words, but not be able to represent what we intended those words to mean. And so I would ask that NCES build in a kind of checks and balance system, such that people who were involved in the four groups here are formally involved in the Penn State process, however that ultimately gets defined.

A fourth issue that came up in our group had to do with what is possible and what is not possible. I was heartened to hear the gentleman from Ford [Ernie Savoie, previous open comment] share his perceptions on that issue. One of the impediments to the kinds of change that will be recommended by our group—and I suspect by all of the other groups as well—in terms of assessment, I think, is the federal government itself. One of the things that I think is real clear is that there needs to be involvement of other agencies besides NCES in this process. Because obviously NCES doesn't have enough money to support the process, in its totality. But it seems to me that NEH—and this is coming out of things that people in our group talked about—and the National Science Foundation to name only a few, should also be involved in this. Because what we're after here has implications for their areas of concern as well. And I know that federal agencies are very territorial beasts, and I would urge, with the new administration, that there be a concerted effort by those in charge of the Department of Education and the sub-units within the Department, to work toward a kind of articulation among agencies with regard to this national assessment of college student learning. Because it reaches beyond the National Center for Education Statistics.

A fifth concern I have is the separation of problem solving, critical thinking, and communications. To some extent that seems to me to be fairly artificial. Partly—and I'll make one brief comment—given the documents I've read on critical thinking and problem solving, it seems to me that it's impossible to assess critical thinking or problem solving without accessing language. And so, in effect, you have an important confound that not only people in critical thinking and problem solving have to be concerned about, but also those of us who are primarily interested in communication. And in terms of antiquity, I would point out that rhetoricians since Plato and Aristotle have been dealing with the very sorts of issues that are now resurfacing in the study of critical thinking and problem solving.

The sixth and last point—and this is something that I will reiterate from the authors' meeting the other night—is that I am not convinced that all of the great things that we want college students to be able to do by way of problem solving, communicating and thinking critically, are necessarily the things that society wants. I think that if we end up, through our educational system, creating a student who is so skilled in all of those areas, that this society, this economic system, this government, and this culture won't know what the hell to do with him. And I really think that's something that needs to be thought about, in the context of developing the national assessment of college student learning, that promises I think to have profound implications for the way instructional activities are carried out, not only in undergraduate colleges but also in professional schools associated with undergraduate schools.

PETER FACIONE, Dean, School of Arts and Sciences, Santa Clara University. (Group 2, reviewer)

Well, you'll be happy to know that I'm prepared to spend several hours talking here at the open mike. The chance to be part of this process, last year and this year, is one for which I'm very grateful. And I've learned a great deal in the process both times. I started thinking about teaching what has come to be called critical thinking—but it wasn't in those days—back in 1967. I took the question of taxonomy seriously back in '82, and got interested in critical thinking assessment in '86, and have done some research that some of you are familiar with since then.

The piece of it that strikes me as somewhat de-emphasized by comparison to its importance at this moment in our development is the piece of it that deals with the personal traits, or characteristics, of people as they approach problems of communication, critical thinking, what have you. We have pretty well identified the skills, the methodologies, and the techniques. We may quibble over the wordsmithing of them, but we have a good idea of what they are, and there even are some instruments out there that attempt to assess them.

But we are just beginning to take the dispositional side of the coin seriously. The dispositional side means things like

- Intellectual curiosity or inquisitiveness, without which people don't even try to solve problems.

- Openmindedness and flexibility, without which people can't see alternatives.

- Systematicity, or the inclination to approach things in an organized, focused and diligent way, without which people don't even persevere, or don't even attempt to form a plan.

- Analyticity, or alertness to opportunities to use critical thinking, without which people kind of stumble along, and don't notice that there are problems that abound all over the place.
- Objectivity, or seeking for the truth, without which people think that problem solving is about only advancing their own interests, rather than taking you where the facts and the reasons lead.
- A level of self-confidence in your own reasoning ability, without which people don't try, and don't make judgments when judgments are called for.
- And certainly a level of cognitive maturity, the recognition that some problems (not unlike the one we've been dealing with the last 2 days) are ill-structured, and sometimes multiple answers are acceptable. And sometimes you have to achieve closure without achieving certainty.

These inclinations need to be investigated, they need to be assessed, they need to be built into what we are doing with the cognitive skills, otherwise the analysis and the inference and the explanation and the evaluation and whatever have we on the skills side won't go anyplace.

In these last 2 days, I think I've learned some important new things, and I'd like to share those with you. I learned that collaboration is an essential aspect of problem solving, if that's indeed what we were engaged in these last couple of days, and we made progress through collaborating. I learned what was reinforced by an earlier speaker, that it takes time to solve interesting problems. And I learned that consensus building is an important part of problem solving. And I learned that we who are in academia have to find ways to communicate effectively with clients in the workplace, or policymakers, who are the people who are going to take the graduates of our institutions into the citizenry and into the workforce. And we who are theoreticians have to be able to connect what we're saying and trying to do with what the people who are the clients are expecting and need.

I'm reminded of two examples of problem solving and critical thinking that I must toss in front of you. Because I, being an academician, tend to approach it in a relatively language-laden way. But I come from a blue-collar background. And I'm reminded of seeing an uncle of mine who was struggling with a problem, because he couldn't get the hammer—which was a certain size—to hit the nail, which was tucked in behind a pipe. And so he cleverly used a pliers, laid the pliers across the head of the nail, and across another piece of material which he put up there as a prop, and hit the pliers which drove the nail. There were no words involved, he just did that. And that strikes me as an incredibly interesting piece of problem solving, although there weren't a lot of words associated with it.

I'm reminded of another example. I was bicycle riding recently, where we had to pass through a fairly narrow gate, with a person approaching from the other side. And the person had a dog on a long leash. The person went to one side of the gate and the dog to the other side of the gate. And it occurred to me at that moment that this person is not disposed to see problems that are coming, and we bike riders are in deep trouble. And so there are many, many contexts, which those two anecdotal comments illustrate, from real life, where the dispositions and the skills have to be used, in fairly instantaneous and nonverbal sorts of ways. And we academics have to find ways to bring that into our analysis.

RICHARD PAUL, Director, Center for Critical Thinking, Sonoma State University.

I'm Richard Paul and, like many of you, I wear a number of hats. I'd like to put on my hat as Chair of the National Council for Excellence in Critical Thinking Instruction (NCECT), and ask you to look with me, for a couple of minutes, at the history of the critical thinking movement. This movement goes back at least 53 years to 1939, to the work of Watson and Glaser, marked by the publication of Glaser's book in 1935, *An Experiment in Critical Thinking,* and the development in 1940 of the *Watson-Glaser Critical Thinking Test.*

The movement has had its ups and downs over the years and for periods of time (1943-58 & 1969-78) had very little impact. Yet throughout the full 53 years there were persons who worked in behalf of critical thinking--Edward Glaser himself being the most notable but many others (e.g., Robert Maynard Hutchins and Mortimer Adler) were also involved--keeping the vision of education based on disciplined reasoning and critical thinking alive.

My involvement dates back about 16 years when I began to work with a small group of faculty members at Sonoma State University to establish the first graduation requirement in critical thinking. At this point, increasing numbers of college professors were coming to recognize that their students did not have the basic intellectual skills and abilities necessary to learn at the level they felt they ought to teach. It became apparent to many that students needed to begin to develop intellectual skills long before they reached college, so some, including myself, began to contact superintendents of schools, to try to interest them in this basic reform. On the whole our efforts were not successful. The typical response was something like, "Yes, yes, very important stuff, but I'm afraid we are not ready for it quite yet. We'll get back to you in time."

Well, in the last 15 years much has transpired to awaken the consciousness of growing numbers of educators to the serious lack of intellectual skills and abilities on the part of our graduates, both high school and college. Fifteen annual international conferences have been held focused explicitly on critical thinking and educational reform. The latest had 1,500 people, 350 sessions, 24 countries and virtually every subject and grade level represented. The question is no longer "Can we get people interested in critical thinking?" It is now "Can we make sure that those who *are interested* understand it in a substantial way, in a way that takes into account the research that has gone into the subject?" The question is now, "Can we make sure that those with a vested interest in influencing education in one way or another, do not make critical thinking into anything whatsoever they choose to call it?" This is the problem we are now facing, the vested and commercial interests are ready to buy critical thinking, package it, and sell it, gutted, of course, of its substance.

Let me give you one vivid example. The California State Department of Education decided to develop a statewide, 8th-grade writing assessment that focused on reasoned evaluation. The students participating were asked to develop an essay in which they assessed--using rationally analyzed criteria, reasons and evidence--the kind of music they like to listen to. They were directed to consider alternative criteria for the evaluation and to justify the criteria they used–again with good reasons. The State Department collected the many thousands of essays, set up teams of assessors, and then, finally, selected just one of the essays for national distribution. When I noticed what they were about–requiring reasoned evaluation–I was of course interested. However, when I read the essay which they were nationally distributing (which they designated as demonstrating "exceptional achievement"), I realized that there was very little reasoning of any kind in the essay and what was there was very poor. Why, you might well ask? Why was a poorly reasoned essay chosen? The problem was that the testing experts at the California State Department of Education could not tell the difference between good reasoning and merely "slick" reasoning–intelligent, witty amusing, articulate; but also illogical and fallacious. When given a choice between good reasoning and "fluff," they went for fluff. Subsequent research by the Center for Critical Thinking has established the fact that this was no fluke, that on average, the majority of teachers give the above poorly reasoned essay about 2 points higher grade than a well-reasoned (but not particularly witty and amusing) one–on a scale of 0-8.

So there is considerable concern among those of us who have spent a good deal of our careers doing research on and arguing for the importance of critical thinking. Our worst nightmare is that of a superficial concept, a nonsubstantial concept of critical thinking introduced into national assessment. Our worst nightmare is having to marshal the full resources of the critical thinking movement into a drive against a national assessment instrument ostensibly claiming, but substantially failing, to involve real critical thinking.

This need not but it may happen. Unfortunately, though most people will freely admit that they are not an expert on this or that, many people who've done no research whatsoever on the subject are ready to identify themselves as instant experts on critical thinking. And so, speaking for the National Council for Excellence in Critical Thinking Instruction, I call upon you to carefully and studiously be on guard against a nonsubstantial substitute for solid critical thinking. I call upon you to make sure that you deeply involve those with incontestible credentials in critical thinking in devising its appropriate assessment. It would be ironic and discouraging if, at the very time critical thinking becomes a national goal, the critical thinking movement has to be marshaled against the very manner in which that goal is being understood, articulated, and assessed. Thank you for your time and attention.

JOHN MUFFO's comments - and some of David Perkins' remarks - have been incorporated into the Group 2 report.

DAVID PERKINS, Harvard University and Project Zero

I have two jobs to do. One is to make a report in behalf of a subgroup of Group 2, and the other is to add a few personal notes. Regarding the first of these, one of the physical products of Group 2 is a kind of expanded taxonomy that included a number of ideas and concerns of the various folks in the group. We felt it was important to frame the expanded taxonomy by a number of ideas or statements that would help to orient Penn State and others who will say it, as to how to take it, how to understand it, and how it might best be used. We also suggest that, in general, the products of all the groups might well be framed by some cautionary and interpretive statements. We think this is important because the products of a group process like this—where you only have a couple of days—inevitably are rather rough-shod, and still contain within them conflicts, ambiguities, and the like.

[See Group 2 report for these remarks and the List of Orienting Points]

Now, if I can ask your indulgence, I'd like to say a few more personal things. Not very many, five of them, and they're short.

First, I think we're at a good historical moment. There is a tide of consciousness concerning higher order thinking, critical thinking, and problem solving. This tide of consciousness emerges from two deep sources. One is the philosophical tradition, which of course is by far the oldest, going back to Plato and Aristotle, and the other is the growth of cognitive psychology and cognitive science, over the past 20 years, an absolute blooming of a science that hardly existed prior to that. Between these two sources there's been enormous clarification, enormous progress and, of course, a good many emergent questions. Nonetheless, we understand a lot, and know a lot that we didn't 20 years ago.

Moreover, this area has passed to some extent into the public consciousness, to some extent into the consciousness of educators, politicians, even the pages of *Time* magazine, on television and the like. Now most of this representation, as we're all too aware, is quite superficial. But nonetheless, some of the words are becoming familiar, some of the labels keep coming up, now and then. And that's very important. I think last night I ventured a quote from Julius Caesar. I'll venture another one, which I think is from Julius Caesar, that goes something like this: "There is a time and tide in the affairs of men"—sorry ladies, but you know, he wrote a while ago—"there is a tide in the affairs of men that, taken at the full, leads on to victory," or something like that. Well, I think we're at the crest of such a tide, and it's part of our responsibility to see that the tide doesn't dissipate.

Second, I think we can speak as well of an emerging consensus. By emerging I mean to imply that there is a lot more that looks consensual now than looked so two or three years ago. For one simple example in the area of critical thinking and problem solving, what Peter mentioned, the importance of dispositions, and indeed the identification of a number of dispositions which people quite generally agree upon. That simply was not the case three years ago. One could point to other things, too. This emerging consensus is part of the tide. I think it's fragile, it needs to be nurtured, and we need to seek more of it. This does not mean there are not conflicts, but there's much to build on.

Third, a little closer to home. I think—in the products contributed to this conference and the previous conference—I would say, creativity is at risk. I don't mean that the products aren't creative, but rather, in my view, that creative thinking may not have received as much attention as it needs. It's present here and there, but perhaps not present enough. I note that this might have particular relevance to—or be appreciated particularly by—industry. One of the paradoxes of the world of industry and business in contrast with the world of academe is that much of the conceptual development, and even the research in the area of creativity, has actually been done in business contexts rather than academic contexts. So both for the sake of that important constituency and in general, I hope that we will look on that area with care, and see that it's well represented.

Fourth, driving the system. One idea about the role of assessment is that if we assess people for what we want, this will tend to drive the system and people will start teaching to what we want. It's a great idea, and an important one, but there's one reason why it tends not to work. The reason it tends not to work is that we sometimes put forth assessments that people don't know how to teach to. When you put forth assessments that people don't know how to teach to, they *don't* drive the system. What they do is to trigger a backlash. This tends to happen, in fact, in the area of critical thinking.

Assessments are posed, teachers—I'm speaking pre-college, here—face these alarming new assessments, and find that they have no idea what to do toward these assessments, find that their students don't do very well, there's a turmoil, and one of two things happens. Either the assessment is removed, in favor of something more conventional, or the assessment is gutted, so what results is really quite superficial. I think that this effort at the college level has to observe the same problem. Several contributors to the conference have noted, and the same in the previous conference, that we need in assembling these assessments to give some thought—not just to how to assess, but—to the kinds of things that teachers might do to prepare students toward those goals. If we don't do that, you can be sure a social process will ensue, whereby the assessments are either cast aside, or gutted.

Finally, CO-thinking. Here's a category I like, and it brings to mind the process of this conference, which is an example of it. This is not a proposed category, but it's an idea I'd like to share with you. CO-thinking is a simple label for the kind of thinking we need a lot of nowadays. CO-thinking is for one thing COoperative thinking. For these kinds of social problems, we need COoperative thinking. It also stands for COmplex thinking. COmplex in a very strong sense, because these kinds of issues are multidimensional, they involve multiple constituencies, multiple perspectives and they are really complex in a horrendous way. Generally, they are much more COmplex than thinking in a particular technical discipline, for example. COmmunicative thinking, because these kinds of problem situations inherently place a large load on our channels of oral and written communication, between one another and the public. And I would add, what may seem a funny one, COmputer and other technology-aided thinking. What do I mean by that? We really can't do this kind of thinking together without computers, without projectors, without flip charts, without magic markers and so forth and so on. We need to use lower and higher technologies to keep track of all the stuff and to organize it. I mention this notion of CO-thinking because it's what we've been doing, and doing tolerably well, but I think it's a whole art unto itself, and if we in our collective effort and other elements in society want to tackle these large social problems, we need to do so through the art and craft of CO-thinking in all its dimensions.

I appreciate the opportunity to participate in this event. I'm sure Shari and Eileen, the three of us being the coauthors of the Group 2 paper, would say the same, but they probably won't come up here and do it, and want to thank you for your attention, and to assure you that in our further work in revising our paper, we'll take what's been said very seriously, indeed.

SHARON RUBIN, Dean, Fulton School of Liberal Arts, Salisbury State University. (Group 4)

I have only two brief comments. In our group we discussed capstone seminars and, less extensively, freshman seminars: both as environments for assessment and as actual processes of assessment. I'd like to alert you to something already available on your campuses, both as an environment and available for the processes of assessment.

Many students on virtually every campus in this country participate in internships, cooperative education, field work, practical and community service learning. And such experiential opportunities are frequently accompanied by academic studies, such as seminars and courses. They offer rich potential, because they are the intersection of the academic critical thinking/communication skills and the work environment; and in many cases, also the public arena in which citizenship takes place. I hope both Penn State and those of you who are interested in assessment on your own campuses will bear this in mind.

A second thing. The question of assessment for what purpose has been discussed quite a lot over the last two days, and I would encourage everybody to keep thinking about that. In our group, we asked a question I think is worthy of being repeated to the larger group. If we don't continually ask the question about how students ethical beliefs and values underlie—or do not underlie—their critical thinking and communication skills, we may be assessing for what Art Levine has called "going first-class on the Titanic."

RICHARD BATTISTONI, Director, Civic Education and Community Service Program, Rutgers University (Group 4)

I'd like to offer a variation of what Sharon just said. I think it's very important that we not confuse—and Steve Witte made this point in our group—means and ends. The Goals statement seems clear: the end is creating better workplaces and creating better citizens. By creating better citizens, I think is meant better communities within the democracy that is America. I want to make sure that in all of these proceedings that we pay attention to an author who isn't here—and wasn't one of the primary authors—but whose work was sent out to all of you, Suzanne Morse.

She wasn't able to come here, but did issue some important statements and concerns I think we all need to think about. When we talk about critical thinking, when we talk about communication, when we talk about reading and writing, the question is "To what purpose? For what?" And when we think about citizenship, we ought to think about communities, about what it means to be a citizen, about what it means to think about not what *I* ought to do, but rather what *we* ought to do. And who that we is comprised of. And I think that's critical, because we could develop the best critical thinkers in terms of certain taxonomies, we could have people who communicate well, we could have people who write and read very well, at a high level or order, but if they don't have a civic framework, if they don't ask the right questions, if they use those skills to advance what they want to do as individuals, or to play a part within society, then I don't know that we're going to create better citizens.

If we just look around at our campuses, they are increasingly diverse. They have diverse people in them. They exist within diverse communities in the larger society as a whole, and yet what I see, when I look around at various campuses, is increased balkanization: people do not collaborate across different contexts. Our institutions of higher education—and that's where I come from—aren't often very good citizens with respect to the communities that surround and nurture them. And I think that we act at our own peril if we fail to keep a civic framework in mind, and if we don't think about citizenship.

Citizenship doesn't just happen. It's not a residual category, and it's not something that happens inadvertently when we create these taxonomies and these skills. Citizenship is a framework, and it's a perspective, and it's something that we always have to keep in common if we're to preserve our democratic institutions. And if we're going to see our young people as actors, as decision makers, not clients of policymakers, not clients of government but actors, in their own right. Corroborators with other fellow citizens in their own right. So, I want to keep that in the forefront of our minds when we think about these things. And I urge you, if you haven't read Suzanne's piece—it's very short—that you read it and also her reviews of the two critical thinking groups as well.

KARL SCHILLING, American Association for Higher Education (Group 3)

My name is Karl Schilling. We talked a lot about resistance in listening in our group, so I would like to say a few things about resistance in talking, too. These caused me to have some pause. I came in, quite frankly, a little skeptical about the utility of the enterprise. I've experienced various phases as we've progressed through this time, but a couple of things concern me.

My primary interest is in assessment work, especially assessment work going on on campus. There's a lot of good work going on throughout the country on campuses, that is very much aimed at improving student learning. It's finally coming of age, I believe, throughout the country, in some very important kinds of ways. My fear is that we will see some—in my view—"bad" money driving out the good money that's occurring on campus, if we adopt a national assessment device which is not going to be aimed (as near as I can tell) at improving student learning. It will drive other assessment work off of campus and will deprive faculty of ownership of the assessment work.

What's been so important in the assessment movement, and particularly what the American Association for Education's forum on assessment has worked very hard at, is ensuring that it stay campus-based, to resist the idea that there would be state controls, state mandates, and state-given. Rather it should be generated from within the campus, because that's where whatever will be designed will actually be used, if it's locally owned. So, that's my first concern, that we might drive out a lot of good assessment work that's been going on.

The second thing is that I do think there is place for the national goals, and that is to create some standards and indicators for campuses to shoot at. I think having some national standards makes a lot of sense, but we ought to allow the institutions the means for assessing those things, and they ought to be required to provide them, in some kind of way. There can be aggregation at the national level. We've moved beyond the very simplistic idea that the only thing you can aggregate is numbers, and that you have to give the same tests. There are some good models for aggregating diverse kinds of data, as long as they're aimed at the same set of indicators. I would urge us to at least think seriously about that model, though I know some of you are resisting that, so it's obviously something that will be debated. But I hope that we will not lose sight of the options: there are other ways to do it. I think that NCES ought to fund, certainly research and development for models of assessment to be right at the campus, but then they should be open to modified change, or develop their own, as part of that.

And finally, if we are going to have it—and I suspect as near as I can tell there is a determination on that point—from my reading of the synopsis of the last conference, and what happened at this conference, it seems to me that most of the advice that was given, by most of the people in attendance, was actually ignored and not really responded to very directly. Because I heard a lot of concerns that this be something based in improvement of the learning of students on campus, as opposed to providing some numbers for people to use to bash institutions with. That's my fear for what might be the first product to come out of here, another set of data and reports that can be used to say that faculty aren't doing good work, and I believe, from the experience that I've had, that faculty on campus are doing extraordinary work, unbelievably good work. A few anecdotes are really driving the system, and I don't want to provide here another tiny bit of data that can be distorted and used against faculty.

But if we are going to have this, let's be sure that we get entry-level data, as well as exit-level data. Because what's inevitable—and Sal said this to me last night—is that, "Yes, the governors will adopt this." And while they're only going to do a sample, he believes the states will drive it in, and every institution in the state will have to do this, and we'll be going with comparative numbers, within a few years. If that happens, we know that whatever we create will have a high correlation with general intelligence factors. All the assessment instruments I've seen have a very high correlation with that and with socioeconomic background and resources.

So, I would urge that the institutions that are doing the hardest work—and the best work—don't get slighted in this. So we would have to go back to a model that we've actually moved away from in assessment, but if we're going to do this, I think we'll have to resurrect it more strongly. This is the value-added kind of approach. Because if we're going to do comparison, that's the only way we're going to get anything that will have any value at all.

DENNIS JONES, National Center for Higher Education Management Systems. (Group 2)

I find it interesting that there's something about being in Group 2 where every individual member of that group has to make individual comments. It was true last year, and it's true this year.

Our organization doesn't pretend to speak for the "client community" in all of this, but we have had a fair amount of experience in dealing with both policymakers and employers of college graduates. I've probably interviewed 800 employers in the last 6 or 8 years and I think I understand a little about what they mean when they say, "We aren't getting what we want in the way of critical thinking, problem solving, and communications"–the kinds of concerns that led to the words in the national goals. As a result of that experience, I'd like to offer a suggestion for a "next step," a step that expands our communication in a very important kind of way.

We've talked about critical thinking a good bit in the last 2 days. I admit I've learned a great deal from my colleagues in this group–about the dimensions of critical thinking, and how complex it really is.

But there's a set of things that we haven't talked about. That is: "As applied to what?" What is the domain of the questions at issue. And what I think is so important as a next step is to find ways, through Delphi or other techniques, to get to the people who are trying to talk to us: to ask them in *their* words, and in their terms, "What is important?" So that we can bring the information back inside and say, "Out of all of this complexity that we're dealing with, how do we bundle all of these various pieces in ways that—when we get all done—they think that they have been served?"

This means, if you will, talking about the dimensions of application, not the dimensions of critical thinking. What is the setting? What is the context? What are the issues that they think are not being addressed well by our higher education institutions? Because uniformly they feel that something is missing. They don't articulate it very well, but I think it's critically important that we find a way to help them talk to us in their language, and to understand both sides of that communication and conversation. It's out of that concern that I think it's also important, as a next step, to ask what evidence will they take, as proof that we have, in fact done our job? What form does that evidence have to be in for them to accept it, regardless of whether or not--in psychometric terms--we have made the case to ourselves. So the questions, "What's the Issue?" and "How do we communicate about it?" These questions must be asked from the user's perspective, whether it's policymakers or employers. I think this is a critical piece, and yet a part we haven't wrestled with very much in the last two days.

Where To Go From Here?

The Workshop, "Assessing Communications and Higher Order Thinking Skills for College Graduates," involved two full days of intense dialogue among dozens of scholars and educators. The preceding remarks were the thoughts of those participants wishing to speak at the joint open session that closed the conference. Another form workshop participants used to assess the conference and their thoughts about the NCES project to develop a national assessment was a survey. They were asked:

(1) What should be the next steps in the process?

(2) What research is needed?

(3) Who should be involved?

(4) What potential problems would you foresee?

In response to the first question, "What should be the next steps in the process?" respondents provided fairly diverse answers, but expressed a consensus in a few critical areas. In particular, that of further development of the taxonomy of skills, abilities, and competencies. Many participants feel that we did not adequately specify the individual competencies at the meeting and that additional clarification and refining are in order. Moreover, in this process, respondents were concerned that the Department and Penn State will fail to consult with them to the degree needed for consensus in the field. Several participants urged the Department and Penn State to retain key members of the work groups (e.g., authors and reviewers) as consultants in attempts to delineate better skills categories. There was also some concern about reconciling the four different assessment groups--problem solving, critical thinking, reading/writing, and listening/speaking--into a holistic assessment design. Most felt it was difficult but important to do this, rather than viewing the four groups as totally separate entities. Respondents hoped for more interaction among experts in these areas in the future. Respondents showed a great deal of interest in, and some concern about, the Penn State Center's Delphi technique, warning the Department and the Center staff against a non-representative sample and urging us to ensure validity and reliability in the instrumentation.

With question two, "What research is needed?", respondents were nearly unanimous in their calls for additional research on ways to link the abilities, skills, and competencies coming from NCES-sponsored deliberations with job and citizenship skill requirements. Several participants note that more and better consultation with business and industry, and also experts in the field of citizenship, could help to bridge existing gaps in these areas. As mentioned under next steps, a unified structure that would embrace the major areas of critical thinking--the problems solving competencies and the several communication skills--is not clearly in sight, and a conceptual methodology may need to be pursued to find it. Methodologically, respondents suggest conducting pilot studies followed by factor analyses for to determine instrument validity (both predictive and content) as well as reliability. And, finally, they urged NCES to consult and examine data from institutions that are currently at the vanguard of the assessment movement; participants mention the University of Tennessee, Knoxville, Rutgers, and Alverno College.

Many of the responses from question two overlapped with question three, "Who should be involved?" Rather than additional representatives from academe, however, most urged additional involvement from business and industry. Many felt that since the business community possesses the ability to assess most directly the outcomes of postsecondary learning, their insights and methods should be factored in at each stage, and of course this constituency is extremely diverse and hard to typify. Moreover, respondents felt that gaining faculty support was absolutely essential to the long-term success of the project, and echoed a major theme from the 1991 workshop, that any assessment's ultimate impact on teaching and learning must go through, and rely on, faculty "buy-in." Some participants also urged that we not neglect the community colleges, especially in light of their general education focus and their transfer function. And several respondents mentioned the involvement of additional professional and academic associations (AERA, AAHE, MLA, etc) and more than a few mentioned the lack of minority representation at the meeting.

Respondents to question four, "Potential problems?", are most concerned about funding for the project, both in terms of the cost of the instrumentation eventually used and the willingness of political operators to allocate funds for sufficiently sophisticated instrumentation. Regarding the instrument itself, some participants were unhappy with the national character of the project, feeling that assessment occurs most naturally at the institutional level; others were afraid that ethnographic methodology would be ignored in favor of an "easy fix" paper and pencil design. There was also a recurring concern that the assessment will differ little from existing measures, and that accountability rather than curricular and instructional improvement is the latent focus of the project. And some respondents were quick to point out that a lack of adequate disciplinary expertise in the instrumentation would result in an invalid instrument, "watered down" to please certain public stakeholders.

The preliminary sets of skills, abilities, and competencies developed by the four work groups, though an important step, may prove to be just an intermediate one. Perhaps more enduring as the discussion continues are the conceptual foundations each author developed on which their skills lists are founded. These can be viewed as a skeleton, around which the ultimate shape of the assessment will grow. But that growth must come in the real world, through conversations about how the proposed skills and assessment methods reflect what is possible and realistic, given the current state of American postsecondary and workplace life. Civics is a set of principles, but citizenship is the very real experience of people wrestling with the complexities of political and social life. Linking these two perspectives has been perhaps the most insistent rallying cry among discussants, surprisingly often from the academics whose theories may be treated roughly in the translation to a politically successful assessment.

1993 Activities

In January 1993, the National Center on Postsecondary Student Teaching, Learning, and Assessment at the Pennsylvania State University. Using information from the 1992 workshop and other sources, through two applications of the "Delphi" technique, the Penn State team created a workable catalogue of skills, abilities, and competencies, based upon the concerns and skill needs of the academic community, workplace, and larger society.[9] The National Center for Higher Education Management Systems (NCHEMS) in 1993, also completed a review of both indirect and proxy measures that are currently available which might be used for the assessment of these skills. Both reports will be published at a later date.

Through past and forthcoming consultations, studies, papers, workshops and publications, NCES is committed to gaining wide acceptance and approval of an assessment framework. Maintaining an ongoing national dialogue about the project is the watchword of the Center's effort. In particular, the 1992 workshop was organized to bridge the gap between general issues and specific student competencies and, in so doing, to provide additional impetus for the evolution of a national assessment process. As with all aspects of this project, the measure of its success depends on the cooperation of the many education and business professionals involved. NCES has a role to facilitate this dialogue and to nourish a consensus, guided, now and in the future, by Objective 5: to enhance the higher order thinking and communication skills of college graduates. NCES embraces the challenge and commits itself to help the nation achieve this end.

---

[9] This Delphi technique involved two applications of a survey to approximately 600 experts and practitioners in the field of postsecondary assessment. Respondents were asked to make suggestions and revisions on an initial set of competencies. After these responses were synthesized, researchers resubmitted the revised set for additional commentary. After both the first and second applications, researchers ranked each skill and competency based on respondents' choices.

☆ U.S. GOVERNMENT PRINTING OFFICE 1994 — 300 - 807 / 23101

ISBN 0-16-045146-9
90000
9 780160 451461

320

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Fourth Class Special
Special Handling

BEST COPY AVAILABLE